MODEL FITTING OF A TWO-FACTOR ARBITRAGE-FREE MODEL FOR THE TERM STRUCTURE OF INTEREST RATES USING MARKOV CHAIN MONTE CARLO

By

Charnchai Leuwattanachotinan

Submitted for the degree of
Doctor of Philosophy
on completion of research in the
Department of Actuarial Mathematics and Statistics,
School of Mathematical and Computer Sciences,
Heriot-Watt University

February 2011

I hereby declare that the work presented in this thesis was carried out by myself at Heriot-Watt University, Edinburgh, except where due acknowledgement is made, and has not been submitted for any other degree.

Charnchai Leuwattanachotinan (Candidate)

Professor Andrew J.G. Cairns (Supervisor)

Doctor George Streftaris (Supervisor)

Date

Abstract

In this thesis we use Markov chain Monte Carlo (MCMC) simulation to calibrate a two-factor arbitrage-free model for the term structure of interest rates which is proposed by Cairns (2004a) based on the positive-interest framework (Flesaker and Hughston, 1996). The model is a time-homogeneous model driven by latent state variables which follow a two-dimensional Ornstein-Uhlenbeck process. A number of MCMC algorithms are developed and employed for estimating both model parameters and latent variables where simulated data are used in the first place in order to validate the algorithms and ensure that they can result in reasonable and reliable estimates before using UK market data. Once the posterior estimates are obtained, we next investigate goodness of fit of the model and eventually assess the impact of parameter uncertainty on the forecasting of yield curves in which the achieved MCMC output can be used directly. Additionally, the developed algorithm is also applied for estimating the two-factor Vasicek term structure model for comparison. We conclude that our algorithms work reasonably well for estimating the Cairns term structure model. The model is then fitted to UK Strips data, and it found to produce reasonable fits for medium- and long-term yields, but we also conclude that some improvement may be required for the short-end of the yield curves.

Acknowledgements

It is truly amazing that I have had opportunities to work with many people over the past couple years of my PhD research. Without their assistance, this thesis definitely could not have been completed and I could not have made it to where I am today. First and foremost, I would like to express my deepest appreciation and gratitude to my supervisor Prof. Andrew Cairns for his invaluable guidance and support for my research. I have learnt a lot from him through our fruitful discussions every time we have a meeting. I also express my appreciation to my second supervisor Dr. George Streftaris who always provides very helpful suggestions on my works and willingly shares his expertise in MCMC to me.

Last, but by no means least, I am very grateful to Dr. Tak Kuen Siu who picked me up to apply for PhD funding from the department. I also would like to acknowledge Prof. Howard Waters and Prof. David Wilkie whom I worked with in my first year and Prof. Zdzislaw Brzeźniak who first introduced me into the stochastic world!

Contents

List of Figures

List of Tables

Introduction

Interest rate plays a substantial role in several kinds of investment. It lends itself to a form of security and also the underlying of many other securities such as derivatives. In light of asset and liability management, a change in interest rates affects the valuation on both sides of a bank's balance sheet. From an economic point of view, the interest rate influences decision making for investors and is a key indicator for the economy that determines the levels of investment, saving and consumption. By definition, interest rate can simply be thought of as the cost of borrowing from one to another, but its modelling is far more complicated than this simple definition suggests.

Typically, interest rates are considered for a wide range of maturities as the yield curve or the term structure. To develop a term structure model, it is necessary to have an insight into the behaviour of interest rates. With reference to Cairns (2004a), the desirable characteristics for a term structure model are as follows:

- All interest rates should be positive and can remain values close to zero.
- The model should be arbitrage-free and framed in continuous-time in order to be able to use for derivative pricing and hedging.
- The model should have a mean-reverting process reflecting that in reality interest rates will not be completely allowed to move freely, but once they reach extreme levels, they will be pulled back to some long-term rates in some way (e.g. an intervention by the Central banks).
- The model should be able to produce the yield curves similar to what we can see in historical data.

In recent decades, a considerable number of arbitrage-free models have been proposed for the term structure of interest rates. At the beginning, much attention was

drawn to one-factor models for the short-term rates or risk-free rates such that the dynamic is characterised by a stochastic process. From empirical research (e.g. Litterman et al., 1991), it is suggested that one-factor models are unlikely to sufficiently capture the dynamics of real market data. Accordingly, multifactor arbitrage-free models have been developed thereafter. Despite an increase of the number of factors and their rigorous frameworks, several multifactor models are yet required to impose some restrictions in order to guarantee interest rates being positive. In effect, this makes those models less flexible and more difficult to implement. Nevertheless, a new framework (the positive-interest framework) introduced by Fleasaker and Hughton (1996), allows us to develop a multifactor model that can ensure the positivity of interest rates in a natural way.

Implementing term structure models, especially multifactor models, is a huge research area. In many cases, the models are incorporated with unobservable state variables so that advanced and modern statistical techniques are necessary for estimation. The main methodologies that frequently appear in the literature related to term structure modelling are maximum likelihood (ML), general method of moments (GMM) and efficient method of moments (EMM) which all follow the frequentist statistical approach. For the Bayesian approach, Markov chain Monte Carlo (MCMC) simulation is the prevailing method for estimation. With application to term structure modelling, Eraker(2001) applied MCMC to fit a two-factor model with a latent stochastic volatility component using weekly US Treasury data from Jannuary 1954 to May 1997. Hu (2005) estimated multifactor affine models using MCMC but those models are not in a general form since Wiener processes are assumed to be uncorrelated ($\rho = 0$). Moreover, the chains of several parameters converge rather poorly and need to be improved. Pooter et al. (2007) employed a Bayesian approach to estimate term structure models incorporating macroeconomic variables but some results appeared to have a substantial problem of convergence since some parameters did not converge at all. Other examples of implementing term structure models using MCMC can also be found in Bester (2004), and Lamoureux and Witte (2002).

In this thesis, we aim at calibrating a specific two-factor arbitrage-free term structure model developed by Cairns (2004a) using UK market data with Markov chain Monte Carlo being the central methodology for our estimation. A new family of

the Cairns models is based on the positive-interest framework that can be used in long-term risk management. In the Cairns model, we are required to estimate both model parameters and time-varying latent state variables, driven by a two-dimensional Ornstein-Uhlenbeck process, and use numerical integration methods to compute the bond prices since their closed-form solution cannot be achieved analytically. All computational work and simulations in this thesis are mainly based on Matlab with some interface to C/C++ where efficiency of programming is essential. In addition, many computations are also carried out in R, S-PLUS and WinBUGS. The thesis is organised as follows.

In Chapter 1, we provide a review of some main term structure models established until recently. Particular attention is given to arbitrage-free models where we also describe the distinct difference between their two main categories, i.e. equilibrium and no-arbitrage models. One-factor models are first presented in order to provide a ground of constructing term structure models under the arbitrage-free principle. Then, we consider multifactor models and focus on the positive-interest framework which is the main framework used for developing the Cairns term structure models.

In Chapter 2, UK market data are investigated. Initially, we describe major interest rate instruments available in the UK and consequently observe their behaviour for short-, medium- and long-term. The key works in this chapter include principal component analysis on the UK data and fitting the Nelson-Siegel model to UK interest rates.

In Chapter 3, we consider modern estimation techniques in term structure modelling. Here we present maximum likelihood (ML), generalised method of moments (GMM) and efficient method of moments (EMM) methods. Some advantages and disadvantages among the methods are provided and we illustrate the GMM and EMM methodologies by estimating the CKLS model (Chan et al., 1992) with 3-month UK Repo rates.

In Chapter 4, we present MCMC methodology which is our core estimation technique for the Cairns term structure model. To begin with, we introduce the Bayes theorem and then describe two main MCMC methods, i.e. the Metropolis-Hastings (MH) algorithm and Gibbs sampler. Next, we use WinBUGS (Spiegelhalter et al., 2003) to estimate a number of time-series models: autoregressive AR(1), bivariate

autoregressive VAR(1) and stochastic volatility models with simulated and market data. After all, we also code the MH algorithm in Matlab to re-estimate the AR(1) model for comparison.

In Chapter 5, we estimate the two-factor Cairns term structure model using MCMC with simulated data where the latent variables are simulated from the exact solution. We first set up the estimation framework and then derive the full posterior distribution of the Cairns bond price. Due to complexity of the model, the MCMC simulation cannot be performed in WinBUGS and therefore we code all algorithms in Matlab with an interface to C++ for computing numerical integrations which is the most time consuming part. Four MCMC algorithms are considered for implementing the model: standard MH algorithm, hybrid MCMC, adaptive MH algorithm and Gibbs sampler with a quadratic approximation. Moreover, we improve the chain convergence by reparameterising the bond posterior distribution and using a blocking strategy.

In Chapter 6, we estimate the two-factor Cairns term structure model using MCMC with monthly UK Strips data. The estimation results are initially discussed and consequently some investigation of goodness of fit of the model with residual analysis are provided.

In Chapter 7, we consider the issue of parameter uncertainty on the forecasting of yield curves using MCMC. The forecasting can be simulated in a straightforward way in which the MCMC output from previous chapter can be directly used. Following Chapter 6, we discuss here the forecasting results on monthly UK Strips data.

In Chapter 8, we fit a two-factor Vasicek term structure model and compare the results to those obtained with the Cairns model. Furthermore, forecast spot rates and annuity values from both models are also compared.

In Chapter 9, we finally conclude with the main results and our contributions and give ideas for further research.

Chapter 1

Term Structure Models of Interest Rates

In this chapter, we review the key models and frameworks for the term structure of interest rates. We first provide basic definitions of several kinds of bonds and interest rates and then discuss the possible factors that might drive the term structure or the yield curve over time. The main term structure models are then described from the descriptive to the arbitrage-free approach and from one-factor to multifactor models. Finally, we pay particular attention to the Cairns term structure models under the positive-interest framework which are the main models for our estimation in subsequent chapters.

1.1 Introduction

1.1.1 Bond Prices and Interest Rates

A bond is a financial obligation that guarantees the buyer to receive a stream of cash flows at specified times in the future. When buying a bond, the buyer is lending money to the bond's issuer and in return will typically be paid nominal interest payments (often referred to coupon payments) and one final payment (face value or principal) at a redemption or maturity date. The coupons are usually determined in terms of rates of return (coupon rates) and conventionally made either annually or semi-annually.

There are many kinds of bonds traded in the markets with a variety of character-

istics. For example, the coupon rates of a bond can be fixed (a fixed-rate bond) or linked to some market rates, e.g. LIBOR rates (a floating-rate bond). If a bond pays no coupon but only a face value at maturity, it will be called a zero-coupon bond. In terms of the issuers, we can also have government bonds, corporate bonds and so on. Given two identical bonds but different issuers, bond prices can be different depending on the creditability or the probability of default on the payments of each issuer.

In a mathematical sense, bond prices and interest rates are interchangeable. In this section, the default-free fixed-rate bonds will be mainly described in greater detail.

1.1.1.1 Zero-Coupon Bonds

A zero-coupon bond is a bond that pays no coupon but a face value of 1 at final time of maturity. Denote $P(t,T)$ as a zero-coupon bond price at time t maturing at time T. A zero-coupon bond is usually sold at discounted price, i.e. $P(t,T) < 1$, for $t < T$, and will have the value at final time T equal to $P(T,T) = 1$. In addition, a zero-coupon bond $P(t,T)$ is also known as a discount factor from time T back to t when calculating the present value of a single cash flow.

1.1.1.2 Spot Rates

Spot rates are agreed interest rates for immediate settlement. The (continuously compounded) spot rate at time t for maturity at time T is defined as

$$R(t,T) = -\frac{\log P(t,T)}{T-t}. \tag{1.1}$$

Thus,

$$P(t,T) = e^{-(T-t)R(t,T)}. \tag{1.2}$$

1.1.1.3 Forward Rates

Forward rates are agreed interest rates over a specified future period of time that often appear in a forward contract. Suppose that a forward contract is constructed at time t, the (continuously compounded) forward rate for time T to S ($t \leq T < S$)

is defined by

$$F(t,T,S) = \frac{1}{S-T}\log\frac{P(t,T)}{P(t,S)}. \tag{1.3}$$

If $T = t$, we will have the forward rate $F(t,t,S)$ equal to the spot rate $R(t,S)$. Moreover, if $S \to T$, such forward rate will be called the instantaneous forward rate which can be written as

$$f(t,T) = \lim_{S\to T} F(t,T,S) = -\frac{\partial}{\partial T}P(t,T). \tag{1.4}$$

Then, it follows that

$$P(t,T) = e^{-\int_t^T f(t,u)du}. \tag{1.5}$$

The instantaneous forward rate $f(t,T)$ can be interpreted as the risk-free interest rate contracted at time t over the infinitesimal time interval from T to $T + dt$.

1.1.1.4 Short Rates

A short rate at time t, denoted $r(t)$, is the instantaneous forward rate at which $T \to t$, i.e.

$$r(t) = f(t,t) = R(t,t). \tag{1.6}$$

The short rate can be regarded as the risk-free interest rate contracted at time t over the infinitesimal time interval from t to $t + dt$.

1.1.1.5 Money-Market Account

A money-market account or cash account is an account that accumulates compound interests of risk-free rates from time to time. It is defined by

$$B(t) = B(0)e^{-\int_0^t r(u)du} \tag{1.7}$$

and hence $dB(t) = r(t)B(t)dt$.

1.1.1.6 Yields to Maturity

A yield to maturity (also known as gross redemption yield) is the constant yield at which the present value of a bond's cash flows equal to the market price. Specifically, suppose that we have a fixed-rate bond with coupons C_i payable annually at time t_i for $i = 1, \ldots, N$ and with face value F payable at t_N. Then, the yield to maturity y is a solution to the equation

$$P(t) = \sum_{i=1}^{N} C_i e^{-y(t_i - t)} + F e^{-y(t_N - t)}, \tag{1.8}$$

where $P(t)$ is the current market price of the bond at time $t \leq t_1$.

1.1.1.7 Par Yields

A par yield is the coupon rate of a new bond that should be priced if it is to be issued at par value (face value). For a fixed-rate bond with coupons C payable annually at time $t + 1, \ldots, t + N = T$ and with face value F payable at T, the par yield at time t for maturity at time T is $\rho(t, T)$ such that

$$\begin{aligned} F &= F\rho(t,T) \sum_{s=t+1}^{T} P(t,s) + FP(t,T) \\ \Rightarrow \quad \rho(t,T) &= \frac{1 - P(t,T)}{\sum_{s=t+1}^{T} P(t,s)}, \end{aligned} \tag{1.9}$$

for $T = t + 1, t + 2, \ldots$.

1.1.2 Term Structure of Interest Rates

Instead of a single rate, interest rates (i.e. zero-coupon bonds, spot rates, forward rates, etc.) are generally considered in terms of the yield curve or the term structure of interest rates. In other words, for a given time, we are interested in observing the behaviour of interest rates for a range of maturities as a whole, particularly for the short-, medium- and long-term. Typically, the yield curves can have a variety of shapes, for instance, upward sloping, downward sloping (inverted), flat, hump-shaped and so on.

There are numerous empirical studies (e.g. Bliss, 1997; Wu, 2003) attempt to ex-

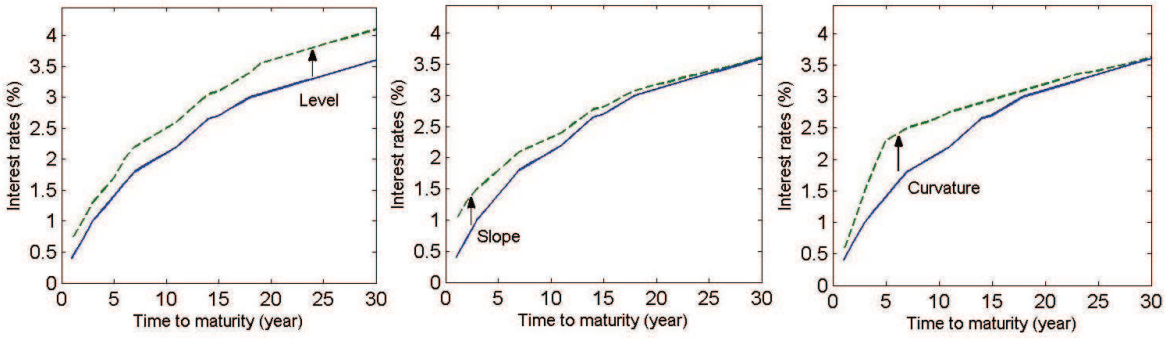

Figure 1.1: Three-factor decomposition of yield curve movement: level (left), slope (middle), and curvature (right).

plain and identify what factors drive changes in the yield curve over time. Recently, much research has gone into unobservable driving variables (Dai and Singleton, 2000) while others (e.g. Diebold et al., 2006) take account into some observable macroeconomic variables such as GDP, inflation and so on. Although, to date, none can perfectly specify all exact factors that characterise the dynamic of the yield curve, it has been agreed that the yield curve movement may be decomposed into three main common factors as so-called "level", "slope" and "curvature" (Litterman and Scheinkman, 1991). As can be seen in Figure 1.1, the effect of "level" makes the yield curve parallelly shift by almost equal magnitude for all maturities, whereas those of "slope" and "curvature" most influence the yield curve for short- and medium-term interest rates respectively. We will see later in the next chapter that this decomposition is also consistent with our results of using the principal component analysis with the UK interest rate data.

1.1.3 The Term Structure Models

Term structure models of interest rates or interest rate models can be separated into two main categories. The first group is the descriptive models that aim at describing the current yield curves by fitting the models to a snapshot of data using statistical curve-fitting techniques. Most models of this kind can result in excellent fitted yield curves but have a little mention of the driving factors or the evolution of term structure of interest rates. Furthermore, they are usually not used for pricing derivatives since the models are not in line with the no-arbitrage principle.

The models in the second category are arbitrage-free models that are widely used

for valuing derivatives and constructing hedging strategies. Of this category, the arbitrage-free models can also be approached by "equilibrium" and "no-arbitrage" principles. The distinct difference between these two approaches is that the equilibrium approach uses the economic reasoning such that today's term structure of interest rates is the output of the process of equilibrating the demand and supply of bonds and other securities in the market (Cox et al., 1985). In turn, the equilibrium models are time-homogeneous in which we might observe that their model parameters do not vary with time. The drawback of this approach is therefore we frequently can find some discrepancy between the observed and theoretical bond prices. For the no-arbitrage approach, the model is designed to perfectly match with today's term structure. Thus, the current bond prices form an input to the model whose term structure will evolve in future time. As a consequence, the no-arbitrage models are typically time-inhomogeneous models that are suitable only for the short-term and which require regular recalibration.

In addition, both kinds of arbitrage-free models can be constructed by either the martingale or the partial differential equations (PDEs) approach. Both methodologies give rise to equivalent results. Some examples can be found in Cairns (2004b) and Heath and Schweizer (2000). A distinct advantage of the PDE approach is that it provides us with a framework for implementation of a model using numerical methods, whereas the martingale approach is claimed to be more mathematically powerful and intuitive. Generally, in order to construct an arbitrage-free model, one is required to:

- define the dynamics of interest rates or latent state variables as stochastic processes under some probability measure (usually starting at the risk-neutral measure),
- solve the stochastic differential equations (SDEs) according to the model setting (hence the difficulty of achieving solutions depends on the model complexity),
- derive or define the market prices of risk and then change the probability measure back to the real world probability after all.

A review of some basic definitions and properties of stochastic differential equations is provided in Appendix A.

1.2 Descriptive Models

Although arbitrage-free models are popular for use of pricing securities and derivatives, they have not much contributed to the yield curve forecasting where the flexibility and parsimony of the model become essentially important. Even worse, some evidence (Duffee, 2002) shows that a simple random walk model may have better forecasts than some equilibrium models such as the affine term structure models. Accordingly, the descriptive models are still often used for the forecasting purpose, especially by the central banks that usually influence the expectation of the short-term interest rates in the market (via the policy bank rate).

Referring to a survey by the Bank for International Settlements (BIS, 2005), it can be found that most major central banks employ either the Nelson and Siegel or the Svensson model (the parametric approach) for estimating zero-coupon yield curves. In case of the Bank of England, however, they have applied the variable roughness penalty (VRP) method (the spline-based approach) which has replaced the previous use of the Svensson model since November 1999. The claimed reason is that for the purpose of assessing monetary conditions, the VRP method is more preferable than the Svensson model as it can better satisfy the criteria of smoothness, flexibility and stability (Anderson and Sleath, 2001).

In this section, we will focus on some main descriptive models by the parametric approach.

1.2.1 Nelson-Siegel Models

Nelson and Siegel (1987) proposed a class of models for instantaneous forward rates in the form of an exponential with a polynomial function. Let $f(t,T)$ be the forward rate at time t maturing at T. Then, the Nelson-Siegel forward rate curve is defined by

$$f(t,T) = \beta_0(t) + (\beta_1(t) + \beta_2(t)\lambda(t)(T-t))e^{-\lambda(t)(T-t)}, \tag{1.10}$$

where $\beta_0(t), \beta_1(t), \beta_2(t)$ and $\lambda(t)$ are the model parameters depending on t.

Despite the advantage of flexibility that there are few parameters in the model, it has been known that the Nelson-Siegel model can produce only one hump in the

yield curve which is inadequate for a more complex shape.

1.2.1.1 Diebold and Li's interpretation

Diebold and Li (2006) interpreted the β parameters in (1.10) as three latent driving factors (level, slope and curvature) consistent with those of Litterman and Scheinkman (1991) described in the earlier section. Specifically, the spot rate yield curve can be written as

$$R(t,T) = L(t) + \left(\frac{1-e^{-\lambda(t)(T-t)}}{\lambda(t)(T-t)}\right)S(t) + \left(\frac{1-e^{-\lambda(t)(T-t)}}{\lambda(t)(T-t)} - e^{-\lambda(t)(T-t)}\right)C(t), \tag{1.11}$$

where $L(t) = \beta_0(t), S(t) = \beta_1(t)$ and $C(t) = \beta_2(t)$ in (1.10).

The loadings on $L(t), S(t)$ and $C(t)$, which are regarded as the "long-term", "short-term" and "medium-term" factors respectively, will be discussed further in next chapter when we estimate the Nelson-Siegel model using UK Strips data. For the model tractability, Diebold and Li suggested to fix $\lambda(t)$ at which the loading term on $C(t)$ is maximised.

1.2.2 Svensson Models

Svensson (1994) extended the Nelson-Siegel model by adding a second exponential term in (1.10) in order to make the forward rate curve be able to generate more complex shapes (e.g. more humps). The Svensson forward rate curve is defined by

$$f(t,T) = \beta_0(t) + (\beta_1(t) + \beta_2(t))(T-t)e^{-\lambda_1(t)(T-t)} + \beta_3(t)(T-t)e^{-\lambda_2(t)(T-t)}. \tag{1.12}$$

1.2.3 Cairns Descriptive Models

The descriptive forward rate curve model proposed by Cairns (1998) is

$$f(t,T) = \beta_0(t) + \sum_{i=1}^{4} \beta_i(t)e^{-\lambda_i(T-t)} \tag{1.13}$$

It can be noticed that, given time t, the model comprises of the constant term β_0 and additional four exponential decay terms.

1.3 One-Factor Arbitrage-Free Models

In an early stage of developing arbitrate-free models, particular attention was given to modelling the short rate (risk-free rate) whose dynamic can be characterised by an Itô process with a stochastic differential equation

$$dr(t) = a(t)dt + b(t)dW(t), \tag{1.14}$$

where $W(t)$ is a Wiener process with respect to a filtration $(\mathcal{F}_t)_{t\geq 0}$ under the real world probability P and $a(t)$ and $b(t)$ are some previsible processes.

In practice, we usually start constructing a short rate model under the risk-neutral probability Q which is equivalent to the real world probability P by making use of the Girsanov theorem and the Radon-Nikodým derivatives (see Appendix A), where the discounted bond price process is a martingale under Q. Once the short rate equation was derived, the zero-coupon bond price $P(t,T)$ can be achieved (under Q) by the Feyman-Kac formula such that

$$P(t,T) = \mathbb{E}_Q\left(e^{-\int_t^T r(u)du}|\mathcal{F}_t\right), \tag{1.15}$$

where $\mathcal{F}_t$ is a filtration generated by a Wiener process $\tilde{W}(t)$ under Q.

In this section, we will consider some well-known one-factor models developed from the dynamic in (1.14) by both equilibrium and no-arbitrage approaches.

1.3.1 Equilibrium Short Rate Models

With the general framework in (1.14), specific one-factor equilibrium models can be developed by defining some particular forms of $a(t)$ and $b(t)$ in which $r(t)$ may also be included but the model parameters are not allowed to be time-varying (hence the models are time-homogeneous).

1.3.1.1 Merton Model

A very first model of this kind is the Merton model (1973) in which the short rate is defined by

$$dr(t) = \alpha dt + \sigma d\tilde{W}(t), \tag{1.16}$$

where $\tilde{W}(t)$ is a Wiener process under the risk-neutral probability Q and α and σ are some positive constants.

Theorem 1.1. ***[Merton Model, 1973]*** *Given* (1.16)*, the zero-coupon bond prices can be achieved at*

$$P(t,T) = \exp[A(t,T) - B(t,T)r(t)], \tag{1.17}$$

where

$$\begin{aligned} B(t,T) &= T-t \\ A(t,T) &= -\frac{1}{2}\alpha(T-t)^2 + \frac{1}{6}\sigma^2(T-t)^3. \end{aligned}$$

As can be seen, the Merton model is simple and hence comes up with several substantial pitfalls. First, the model does not have a mean-reverting property in which it is commonly known (Cairns, 2004a) that in reality interest rates will not drift freely but at some point they will eventually be pulled back to some long-term rates. Second, interest rates generated by the Merton model can become negative which is unrealistic.

1.3.1.2 Vasicek Model

The Vasicek model (1977) (also known as the Ornstein-Uhlenbeck process) defines the dynamic of short rate as

$$dr(t) = \alpha(\mu - r(t))dt + \sigma d\tilde{W}(t), \tag{1.18}$$

where $\tilde{W}(t)$ is a Wiener process under the risk-neutral probability Q and α, μ and σ are strictly positive constants.

Theorem 1.2. ***[Vasicek Model, 1977]*** *Given* (1.18)*, we have the zero-coupon*

bond prices

$$P(t,T) = \exp[A(t,T) - B(t,T)r(t)], \tag{1.19}$$

where

$$\begin{aligned} B(t,T) &= \frac{1-e^{-\alpha(T-t)}}{\alpha} \\ A(t,T) &= (B(t,T)-(T-t))\left(\mu - \frac{\sigma^2}{2\alpha^2}\right) - \frac{\sigma^2}{4\alpha}B(t,T)^2. \end{aligned}$$

The Vasicek model is improved from the Merton model by incorporating a mean-reverting process in the drift term but unfortunately interest rates still can become negative.

1.3.1.3 Cox-Ingersoll-Ross Model

Cox, Ingersoll and Ross (CIR, 1985) proposed the short rate model that can prevent negative interest rates by having the square-root process $\sqrt{r(t)}$ in the diffusion coefficient of the Vasicek model. Namely,

$$dr(t) = \alpha(\mu - r(t))dt + \sigma\sqrt{r(t)}d\tilde{W}(t), \tag{1.20}$$

where $\tilde{W}(t)$ is a Wiener process under the risk-neutral probability Q and α, μ and σ are strictly positive constants.

Theorem 1.3. ***[CIR Model, 1985]*** *Given* (1.20)*, the zero-coupon bond prices are*

$$P(t,T) = \exp[A(t,T) - B(t,T)r(t)], \tag{1.21}$$

where

$$\begin{aligned} A(t,T) &= \frac{2\alpha\mu}{\sigma^2}\log\left(\frac{2\gamma e^{(\gamma+\alpha)(T-t)/2}}{(\gamma+\alpha)(e^{\gamma(T-t)}-1+2\gamma}\right), \\ B(t,T) &= \frac{2(e^{\gamma(T-t)}-1)}{(\gamma+\alpha)(e^{\gamma(T-t)}-1)+2\gamma}, \\ \gamma &= \sqrt{\alpha^2+2\sigma^2}. \end{aligned}$$

1.3.1.4 Chan-Karolyi-Longstaff-Sanders Model

Chan, Karolyi, Longstaff and Sanders (CKLS, 1992) generalised the CIR model in such a way that, rather than fixing at 0.5, the power of $r(t)$ in the diffusion term is treated as an additional parameter, i.e.

$$dr(t) = \alpha(\mu - r(t))dt + \sigma r(t)^{\gamma} d\tilde{W}(t), \tag{1.22}$$

where $\tilde{W}(t)$ is a Wiener process under the risk-neutral probability Q and α, μ, σ and γ are some positive constants.

From their empirical study, the estimate of γ is approximately equal to 1.5 from fitting the model using the generalised method of moments (GMM) to the historical one-month US Treasury Bill yields from June 1964 to December 1989. Apart from the special cases of $\gamma = 0$ (Vasicek) and $\gamma = 0.5$ (CIR), there are no analytical expressions for bond prices.

1.3.2 No-Arbitrage Short Rate Models

Under the framework (1.14), a no-arbitrage short rate model can also be developed in which we frequently can find some parameters in the previsible processes be time-varying. As a consequence, the model is time-inhomogeneous and requires an initial term structure for computing interest rates.

1.3.2.1 Ho-Lee Model

Ho and Lee (1986) generalised the Merton model by setting the short rate as

$$dr(t) = \alpha(t)dt + \sigma d\tilde{W}(t), \tag{1.23}$$

where $\tilde{W}(t)$ is a Wiener process under the risk-neutral probability Q, $\alpha(t)$ is a deterministic function and σ is a positive constant.

Theorem 1.4. ***[Ho-Lee Model, 1986]*** *Suppose that the dynamic of short rates follows the SDE in* (1.23)*. By the Fayman-Kac formula and with*

$$\alpha(T) = \frac{\partial}{\partial T} f(0,T) + \sigma^2 T, \tag{1.24}$$

where $f(0,T) = -\partial P(0,T)/\partial T$ *is the initial forward rates, it can be proved that*

$$P(t,T) = \exp[A(t,T) - B(t,T)r(t)], \tag{1.25}$$

where

$$\begin{aligned} B(t,T) &= T - t, \\ A(t,T) &= \log \frac{P(0,T)}{P(0,t)} + B(t,T)f(0,t) - \frac{1}{2}\sigma^2 t B(t,T)^2. \end{aligned}$$

Hence,

$$r(t) = f(0,t) + \frac{1}{2}\sigma^2 t^2 + \sigma\tilde{W}(t), \tag{1.26}$$

where $f(0,t) = r(0)$.

1.3.2.2 Hull-White Model

Hull and White (1990) proposed the constant parameter μ in the Vasicek model as a deterministic function depending on time. That is,

$$dr(t) = \alpha(\mu(t) - r(t))dt + \sigma d\tilde{W}(t), \tag{1.27}$$

where $\tilde{W}(t)$ is a Wiener process under the risk-neutral probability Q, $\mu(t)$ is a deterministic function and α and σ are some positive constants.

Theorem 1.5. ***[Hull-White Model, 1990]*** *Given the dynamic of short rates as the SDE in* (1.27)*, by the Fayman-Kac formula and with*

$$\mu(T) = \frac{1}{\alpha}\frac{\partial}{\partial T}f(0,T) + f(0,T) + \frac{\sigma^2}{2\alpha^2}(1 - e^{-2\alpha T}), \tag{1.28}$$

where $f(0,T)$ *is the initial forward rates, it can be shown that*

$$P(t,T) = \exp[A(t,T) - B(t,T)r(t)], \tag{1.29}$$

where

$$B(t,T) = \frac{1 - e^{-\alpha(T-t)}}{\alpha},$$

$$A(t,T) \;=\; \log\frac{P(0,T)}{P(0,t)} + B(t,T)f(0,t) - \frac{\sigma^2}{4\alpha}(1-e^{-2\alpha t})B(t,T)^2.$$

Accordingly,

$$r(t) = f(0,t) + \frac{\sigma^2}{2\alpha^2}(1-e^{-\alpha t})^2 + \sigma\int_0^t e^{-\alpha(t-s)}d\tilde{W}(s). \tag{1.30}$$

1.4 Multifactor Arbitrage-Free Models

In the previous section, we considered several arbitrage-free models which are the one-factor models. In other words, there is only one underlying source of randomness (i.e. one Wiener process) incorporated in those models. As already mentioned in the introduction, it has been known that one-factor model is unlikely to capture all features of interest rates and therefore multifactor models have been developed.

1.4.1 Affine Term Structure Models (ATSMs)

In fact, all the one-factor short rate models previously described have the bond prices in an affine form of $\exp[A(t,T) - B(t,T)r(t)]$, where $A(t,T)$ and $B(t,T)$ are the specific functions. More precisely, those models belong to a class of the affine term structure models (ATSMs) proposed by Duffie and Kan (1996).

Under the ASTM framework, the short rate is specified as an affine function of the state variables $X(t) = (X_1(t), X_2(t), ..., X_N(t))'$, i.e.

$$r(t) = \delta_0 + \sum_{i=1}^{N}\delta_i X_i(t) \equiv \delta_0 + \delta_x' X(t), \tag{1.31}$$

where $\delta_x = (\delta_0, \ldots, \delta_N)'$ is a vector of constants and the dynamic of state variables X(t) is defined by the SDE

$$dX(t) = K(\theta - X(t)) + \Sigma F(X(t))d\tilde{W}(t), \tag{1.32}$$

where $\tilde{W}(t)$ is an N-dimensional independent Wiener process under the risk-neutral probability Q, θ is a constant $N \times 1$ vector, K and Σ are constant $N \times N$ matrices

and $F(X(t))$ is the diagonal $N \times N$ matrix such that

$$F(X(t)) = \begin{bmatrix} \sqrt{\alpha_1 + \beta_1' X(t)} & 0 & \cdots & 0 \\ 0 & \sqrt{\alpha_2 + \beta_2' X(t)} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\alpha_N + \beta_N' X(t)} \end{bmatrix}, \tag{1.33}$$

where $\alpha_1, \ldots, \alpha_N$ are constants and each $\beta_i = (\beta_{i1}, \ldots, \beta_{iN})'$ is a constant vector.

Theorem 1.6. ***[The ATSM Framework, 1996]*** *Providing the general framework for the short rate from* (1.31) *to* (1.33)*, the zero-coupon bond prices at time t maturing at T have the affine form*

$$P(t,T) = \exp[A(t,T) - B'(t,T)X(t)], \tag{1.34}$$

where $A(t,T)$ and $B(t,T)$ satisfy the ordinary differential equations (ODEs)

$$\begin{aligned} \frac{dA(t,T)}{dT} &= -\theta' K B(t,T) + \frac{1}{2}\sum_{i=1}^{N} [\Sigma' B(t,T)]_i^2 \alpha_i - \delta_0, \\ \frac{dB(t,T)}{dT} &= -K' B(t,T) - \frac{1}{2}\sum_{i=1}^{N} [\Sigma' B(t,T)]_i^2 \beta_i - \delta_x. \end{aligned}$$

with the initial conditions $A(0) = 0$ and $B(0) = 0_{N\times 1}$.

1.4.1.1 A Canonical Form of the ATSMs

It can be noticed from the ATSM framework that not all sets of the parameter vector $\Theta = (K, \theta, \Sigma, B, \alpha)$, where $B = (\beta_1, \ldots, \beta_N)$, give rise to positive interest rates. Specifically, we need to impose some restrictions on Θ in order to ensure that the terms $\alpha_i + \beta_i X(t)$ of the conditional variances in (1.33) are strictly positive for all i (such specification for Θ will be referred as "admissible").

Dai and Singleton (2000) provided an excellent specification analysis regarding this. They define the canonical representation of the admissible N-factor ATSMs as the following definition.

Definition 1.7. [**A Canonical Form of Admissible N-factor ATSMs**] Let $m = rank(B)$ be an index for the dependency degree of the conditional variances

$F(X(t))$ on the number of the latent variables $X(t)$. Further, for each m, $X(t)$ is partitioned as $X' = (X^{U\prime}, X^{L\prime})$, where X^U and X^L are $m \times 1$ and $(N-m) \times 1$ vectors respectively. Then, a canonical form of admissible N-factor ATSMs can be defined by the specification of the parameter vector Θ such that

$$K = \begin{pmatrix} K^{UU}_{m\times m} & 0_{m\times(N-m)} \\ K^{LU}_{(N-m)\times m} & K^{LL}_{(N-m)\times(N-m)} \end{pmatrix},$$

for $m > 0$, and K is either upper or lower triangular matrix for $m = 0$,

$$\theta = \begin{pmatrix} \theta^U_{m\times 1} \\ 0_{(N-m)\times 1} \end{pmatrix},$$

$$\Sigma = I,$$

$$\alpha = \begin{pmatrix} 0_{m\times 1} \\ 1_{(N-m)\times 1} \end{pmatrix},$$

$$B = \begin{pmatrix} I_{m\times m} & B^{UL}_{m\times(N-m)} \\ 0_{(N-m)\times m} & 0_{(N-m)\times(N-m)} \end{pmatrix},$$

with the parameter restrictions:

$$\delta_i \geq 0, \qquad m+1 \leq i \leq N,$$

$$K_i\theta = \sum_{j=1}^{m} K_{ij}\theta_j > 0, \qquad 1 \leq i \leq m,$$

$$K_{ij} \leq 0, \qquad 1 \leq i \leq m, \quad j \neq i,$$

$$\theta_i \geq 0, \qquad 1 \leq i \leq m,$$

$$B_{ij} \geq 0, \qquad 1 \leq i \leq m, \quad m+1 \leq j \leq N.$$

1.4.2 Quadratic Term Structure Models (QTSMs)

In place of an affine function, Ahn, Dittmar and Gallant (2002) introduced the short rate in a quadratic form of the state variables $X(t) = (X_1(t), X_2(t), ..., X_N(t))'$ such

that

$$r(t) = \alpha + \beta' X(t) + X(t)' \Psi X(t), \tag{1.35}$$

where α is a constant, β is an $N \times 1$ constant vector and Ψ is an $N \times N$ constant matrix. In addition, the dynamic of state variables $X(t)$ is governed by the SDE

$$dX(t) = (\mu + \xi X(t))dt + \Sigma dW(t), \tag{1.36}$$

where $W(t)$ is an N-dimensional Wiener process under some physical probability measure P, μ is a constant $N \times 1$ vector and ξ and Σ are constant $N \times N$ matrices.

It is worth noting that the QTSM framework initially formulates the models under the probability measure P rather than Q. Nevertheless, by the Girsanov theorem and defining $d\tilde{W}(t) = dW(t) + \Sigma^{-1}(\delta_0 + \delta_1(X(t))dt$, the dynamic of the state variables under Q can be

$$dX(t) = [\mu - \delta_0 + (\xi - \delta_1)X(t)]dt + \Sigma d\tilde{W}(t), \tag{1.37}$$

where $\tilde{W}(t)$ is an N-dimensional Wiener process under the risk-neutral measure Q, μ and δ_0 are constant $N \times 1$ vectors and ξ, δ_1 and Σ are constant $N \times N$ matrices.

Theorem 1.8. ***[The QTSM Framework, 2002]*** *Given the dynamic of the short rate in* (1.35) *and the state variables in* (1.37)*, the zero-coupon bond prices at time t maturing at T have the quadratic form*

$$P(t,T) = \exp[A(t,T) + B'(t,T)X(t) + X'(t)C(t,T)X(t)], \tag{1.38}$$

where $A(t,T), B(t,T)$ and $C(t,T)$ satisfy the ordinary differential equations (ODEs)

$$\begin{aligned}
\frac{dC(t,T)}{dT} &= 2C(t,T)\Sigma\Sigma' C(t,T) + (C(t,T)(\xi - \delta_1) + (\xi - \delta_1)' C(t,T)) - \Psi, \\
\frac{dB(t,T)}{d\tau} &= 2C(t,T)\Sigma\Sigma' B(t,T) + (\xi - \delta_1)' B(t,T) + 2C(t,T)(\mu - \delta_0) - \beta, \\
\frac{dA(t,T)}{d\tau} &= tr[\Sigma\Sigma' C(t,T)] + \frac{1}{2} B(t,T)' \Sigma\Sigma' B(t,T) + B'(t,T)(\mu - \delta_0) - \alpha.
\end{aligned}$$

with the initial conditions $A(0) = 0, B(0) = 0_{N\times 1}$ *and* $C(0) = 0_{N\times N}$.

1.4.2.1 A Canonical Form of the QTSMs

Comparing to the ATSMs, the restrictions for ensuring the positive interest rates to the QTSMs are much simpler due to the existence of a homoscedastic diffusion matrix Ψ for the state variables. Specifically, the canonical form of the QTSMs satisfies

- Ψ is a positive semidefinite matrix such that

$$\Psi = \begin{bmatrix} 1 & \Psi_{12} & \cdots & \Psi_{1N} \\ \Psi_{12} & 1 & \cdots & \Psi_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ \Psi_{1N} & \Psi_{2N} & \cdots & 1 \end{bmatrix},$$

- $\alpha, \mu \geq 0$ and $\beta = 0_{N\times 1}$,
- ξ and δ_1 are lower triangular matrices,
- Σ is a diagonal matrix.

1.4.3 Heath-Jarrow-Morton Models

The new framework introduced by Heath, Jarrow and Morton (HJM, 1992) made a huge impact on modelling the term structure of interest rates. Instead of focusing on the short rates, they directly specify the dynamic of the instantaneous forward rates as the SDE

$$df(t,T) = \alpha(t,T)dt + \sigma(t,T)'d\tilde{W}(t), \tag{1.39}$$

for all $0 \leq t \leq T$, where $\tilde{W}(t)$ is an N-dimensional Wiener process under the risk-neutral measure Q, $\alpha(t,T)$ is a drift process and $\sigma(t,T) = (\sigma_1(t,T), \ldots, \sigma_N(t,T))'$ is a volatility process. Under the real world measure P, one can define $dW(t) = d\tilde{W}(t) + \gamma(t)dt$, where $W(t)$ is an N-dimensional Wiener process under P and $\gamma(t)$ is the market price of risk.

Theorem 1.9. ***[The HJM Framework, 1992]*** *The forward rate models developed under the SDE in* (1.39) *are arbitrage-free if and only if*

$$\alpha(t,T) = -\sigma(t,T)'S(t,T), \tag{1.40}$$

$$\sigma_i(t,T) = -\frac{\partial}{\partial T}S_i(t,T), \tag{1.41}$$

where $S(t,T) = (S_1(t,T),\ldots,S_N(t,T))'$ *is a previsible process. Hence, we also have the short rates*

$$r(t) = f(t,t) = f(0,t) - \int_0^t \sigma(s,t)'S(s,t)ds + \int_0^t \sigma(s,t)'d\tilde{W}(s) \tag{1.42}$$

and the dynamic of zero-coupon bond prices

$$dP(t,T) = P(t,T)[r(t)dt + S(t,T)'d\tilde{W}(t)]. \tag{1.43}$$

The HJM framework allows us to develop specific arbitrage-free forward rate models through the volatility term structure $\sigma(t,T)$. By choosing an appropriate volatility function, it is possible to reduce the HJM model to simpler models such as the Ho-Lee and Hull-White models.

The popular model under the HJM framework is the LIBOR market model developed by Brace, Gatarek and Musiela (BGM, 1997) where the volatility structure is derived by using the LIBOR rate process. Rather than forward rates, the model can directly use the LIBOR rates which are simple interest rates genuinely quoted in the markets. For the BGM model, LIBOR rates are assumed to follow the log-normal distribution in which, contrary to the HJM model, the explosion can be avoided. Furthermore, the BGM model also allows us to develop a straightforward framework for pricing interest rate derivatives such as caps, floors and swaptions (generally, there are no simple formulae for pricing such derivatives under the HJM models).

1.5 Positive-Interest Framework

We can see that the term structure models previously discussed require some restrictions to ensure interest rates be positive. To an extent, this can be thought of a

drawback that, possibly, limits some favourable characteristics of the models. To overcome this, Flesaker and Hughston (FH, 1996) introduced the positive-interest framework that can guarantee the positivity of interest rates in a natural and constructive way.

Proposition 1.10. ***[The FH Zero-Coupon Bond Price]*** *Assume that $M(t,s)$ is a family of strictly positive martingales with respect to a filtration $(\mathcal{F}_t)_{t\geq 0}$ under $\hat{P}$, where $M(0,s)=1$ for all s. Also suppose that a zero-coupon bond price $P(t,T)$ is differentiable at T and $\phi(s)$ is the strictly positive deterministic function such that*

$$\phi(s) = \frac{\partial}{\partial s} P(0,s). \tag{1.44}$$

Then, the FH zero-coupon bond price, defined by

$$P(t,T) = \frac{\int_T^\infty \phi(s) M(t,s) ds}{\int_t^\infty \phi(s) M(t,s) ds}, \tag{1.45}$$

is arbitrage-free and result in positive interest rates almost surely. Moreover, we also have the forward rate

$$f(t,T) = -\frac{\partial}{\partial T} \log P(t,T) = \frac{M(t,T)\phi(T)}{\int_T^\infty M(t,s)\phi(s) ds} \tag{1.46}$$

and the short rate

$$r(t) = f(t,t) = \frac{M(t,t)\phi(t)}{\int_t^\infty M(t,s)\phi(s) ds}. \tag{1.47}$$

A more general form of the FH framework was subsequently proposed by Rutkowski (1997) and Roger (1997) as the following theorem.

Theorem 1.11. *Suppose that $(\Omega, \mathcal{F}, \hat{P})$ is a probability space and $\hat{W}(t)$ is a finite Wiener process with respect to a filtration $(\mathcal{F}_t)_{t\geq 0}$ under some pricing measure $\hat{P}$. Let $P(t,T)$ be a zero-coupon bond price at time t maturing at T and $A(t)$ be a strictly positive supermartingale under $\hat{P}$. Then, arbitrage-free zero-coupon bond prices can be defined by*

$$P(t,T) = \frac{\mathbb{E}_{\hat{P}}[A(T)|\mathcal{F}_t]}{A(t)} \tag{1.48}$$

and hence the corresponding interest rates also remain positive almost surely.

Proof. See Cairns (2004b). □

It is noted that the FH form can be achieved by setting $A(t) = \int_t^\infty M(t,s)\phi(s)ds$ in (1.48).

1.5.1 Arbitrage-Free Cairns Term Structure Models

By specifying some process $M(t,s)$ and deterministic function $\phi(s)$ in the FH bond price in (1.45), Cairns (2004a) developed a specific multifactor term structure model which are arbitrage-free and can be used for long-term risk management. The model is Markov and time-homogeneous where the state variables follow standard, correlated Ornstein-Uhlenbeck processes. In particular, when the parameter and state variable values are given appropriately, the model can produce a yield curve similar to those of the Japanese yield curves such that they remain flat and very low just above zero.

Theorem 1.12. ***[The Arbitrage-Free Cairns Multifactor Models]*** *Assume that $(\Omega, \mathcal{F}, \hat{P})$ is a probability space and $\hat{W}_1(t), \ldots, \hat{W}_n(t)$ are n independent Wiener processes with respect to a filtration $(\mathcal{F}_t)_{t\geq 0}$. Next, define $M(t,T)$ as the family of martingales such that*

$$\begin{aligned} M(0,T) &= 1, \text{ for all } T, \\ dM(t,T) &= M(t,T)\sigma(t,T)'d\hat{Y}(t) = M(t,T)\sigma(t,T)'Cd\hat{W}(t) \\ &= M(t,T)\sum_{i=1}^{n}\sigma_i(t,T)d\hat{Y}(t), \end{aligned} \tag{1.49}$$

where $d\hat{Y}(t) = Cd\hat{W}(t)$, $\hat{Y}(0) = 0$, and the matrix C is the Cholesky matrix such that $CC' = (\rho_{ij})_{i,j=1}^n$ (an instantaneous correlation matrix). Suppose also that $\sigma_i(t,T) = \sigma_i \exp[-\alpha_i(T-t)]$ and

$$\phi(s) = \phi \exp\left[-\beta s + \sum_{i=1}^{n}\sigma_i\hat{x}_i e^{-\alpha_i s} - \frac{1}{2}\sum_{i,j=1}^{n}\frac{\rho_{ij}\sigma_i\sigma_j}{\alpha_i+\alpha_j}e^{-(\alpha_i+\alpha_j)s}\right]. \tag{1.50}$$

By making use of the FH framework, the Cairns multifactor model for zero-coupon bond prices at time t maturing at time T can be achieved at

$$P(t,T) = \frac{\int_{T-t}^{\infty} H(u, X(t))du}{\int_0^\infty H(u, X(t))}, \tag{1.51}$$

where $X(t)$ follows an n-dimensional Ornstein-Uhlenbeck process; $dX_i(t) = -\alpha_i X_i(t)dt + d\hat{Y}(t)$, for $i = 1, \ldots, n$, and

$$H(u,x) = \exp\Big[-\beta u + \sum_{i=1}^{n} \sigma_i x_i \, e^{-\alpha_i u} - \frac{1}{2} \sum_{i,j=1}^{n} \frac{\rho_{ij}\sigma_i\sigma_j}{\alpha_i + \alpha_j} e^{-(\alpha_i+\alpha_j)u} \Big]. \tag{1.52}$$

Furthermore, we also obtain the forward rate

$$f(t,T) = \frac{H(T-t, X(t))}{\int_{T-t}^{\infty} H(u, X(t))du} \tag{1.53}$$

and the short rate

$$r(t) = \frac{H(0, X(t))}{\int_{0}^{\infty} H(u, X(t))du}. \tag{1.54}$$

Proof. See Cairns (2004a). □

1.5.1.1 Remarks on the Arbitrage-Free Cairns Term Structure Models

- Whereas the original FH framework allows us to develop a no-arbitrage model such that the current market prices can be used to form an input into the model through the deterministic function $\phi(s)$ in (1.44), the specific Cairns term structure models turn to be a kind of arbitrage-free models which are time-homogeneous and Markov (i.e., the state variables $X(t)$ are Markov). Accordingly, the differences between theoretical and observed market prices may be observed. Nevertheless, it can be argued that the market friction (such as transaction costs and so on) may prevent such arbitrage opportunities to some extent if the number of factors n in the model is large enough (at least $n \geq 2$).

- Although a numerical integration is required for obtaining a bond price from (1.51), it can be done quickly and accurately since we are dealing only with a one-dimensional integral.

- For the one-factor Cairns model, it can be proved that if the short rate $r(t)$ goes small, the model can behave like the Black and Karansinski (1991) model and like the Vasicek model when $r(t)$ gets large.

- The model is first constructed in some pricing measure $\hat{P}$ in which the state

variables are multivariate normal distributed. To change the measure back to the real world probability P with preserving the normality, the market prices of risk are structured in such a way that

$$\begin{aligned} d\tilde{W}(t) &= d\hat{W}(t) - V(t,t)dt \\ dW(t) &= d\tilde{W}(t) + \gamma(t)dt \\ \Rightarrow dW(t) &= d\hat{W}(t) + \theta dt, \end{aligned}$$

where $\hat{W}(t), \tilde{W}(t)$, and $W(t)$ are Wiener processes under the pricing measure $\hat{P}$, the risk-neutral Q and the real world probability P respectively, and $\gamma(t) = V(t,t) + \theta$ for some constant vector θ where $V(t,t)$ is some $n \times 1$ vector process (See Cairns, 2004). Thus, the dynamic of the latent state variable $X_i(t)$ under P will be

$$dX_i(t) = \alpha_i(\gamma_i - X_i(t))dt + \sum_{j=1}^{n} c_{ij} dW_i(t),$$

where $\gamma_i = -\alpha_i^{-1} \sum_{j=1}^{n} c_{ij}\theta_j$ and c_{ij} are the elements of the Cholesky matrix C.

Chapter 2

An Analysis of UK Interest Rates

This chapter is mainly devoted to explore and analyse UK interest rate data. To begin with, we describe major interest rate instruments in the UK market and then particularly investigate the movement behaviour of the short-, medium- and long-term UK interest rates. Consequently, we carry out two further analyses. First, we fit the Nelson-Siegel model using ordinary least squares (OLS) with monthly UK Strips in order to observe the dynamic of the three latent driving factors (level, slope and curvature) as described by Diebold and Li (2006). Second, we conduct the principal component analysis (PCA) to identify main principal components of the UK data and subsequently investigate the standardised residuals after utilising the PCA results.

2.1 UK Interest Rate Markets and Instruments

In the UK, interest rates are generally facilitated in two main markets: the gilt and money markets. Gilts are the UK government bonds which have been managed under the UK Debt Management Office (DMO) (Before 1 April 1998, it was the responsibility of the Bank of England) whereas most major money market instruments (e.g. LIBOR and Repo rates) are currently handled by the British Bankers' Association (BBA).

2.1.1 UK Gilt Market

There are three main different categories of securities in the gilt market: conventional gilts (coupon-bearing bonds), index-linked gilts and gilt Strips (zero-coupon bonds).

It should be mentioned that, before 1996, the capital gains were taxed with different rates for the low- and high-coupon gilts. As a result, the market yields at that time may not reflect the true returns. Specifically, yields on the high-coupon gilts tended to be higher than those on the low-coupon ones due to tax asymmetries. Therefore, one should bear this in mind when considering the UK historical gilt data.

2.1.1.1 Conventional Gilts

Conventional gilts are the simplest form of UK government bonds. The DMO has concentrated recently on issuing conventional gilts for a certain maturities of around 5, 10, 30, 40 and 50 years with the semi-annual coupon payments on specific aligned coupon dates: 7 March/7 September and 7 June/7 December. The prices of conventional gilts are quoted in terms of £100 nominal.

2.1.1.2 Index-Linked Gilts

Index-linked gilts are the conventional gilts that their coupon payments and the principal are adjusted by the UK Retail Prices Index (RPI). As a consequence, the coupons on index-linked gilts reflect the real yields (rather than the nominal yields). Since the RPI data is usually not promptly available for the date of valuation, the payment on an index-linked gilt will be calculated relying on lagged RPI data in which all new index-linked gilts in the UK have used the three-month lag since September 2005 (in the past, there was also eight-month lag index-linked gilts).

2.1.1.3 Gilt Strips

Strips (Separate Trading of Registered Interest and Principal Securities) are zero-coupon bonds constructed by breaking down a conventional gilt into individual cash flows (as so-called "stripping" process). For instance, a five-year conventional gilt could be stripped into 11 zero-coupon bonds: one principal and ten semi-annual coupon payments.

Strips have been traded in the UK since December 1997 with the same specific aligned coupon dates as the conventional gilts (7 March, June, September and December). Referring to the DMO formulae for calculating gilt prices from yields (DMO,

2005), the relationship between Strips price and yield is defined by

$$P = \frac{100}{(1 + \frac{y}{2})^{\frac{r}{s}+n}}, \tag{2.1}$$

where

P = Price per £100 nominal of Strip,
y = Gross redemption yield (in decimal),
r = Number of calendar days from the settlement to the next quasi-coupon date,
s = Number of calendar days in the quasi-coupon period in which the settlement date occurs,
n = Number of remaining full quasi-coupon periods after the current period.

NB: r and s are not adjusted for non-working days.

Example 2.1. *Consider one Strip maturing on 7 June 2015. If the settlement is made on 26 June 2010 at which the current market price is* 85.55 *pounds. Then, we first know that the previous and the next quasi-coupon dates of this Strip are 7 June 2010 and 7 December 2010 respectively. Next, one can count the days and obtain* $r = 166$ *days,* $s = 183$ *days and* $n = 9$ *periods. Accordingly, yield can be computed from* (2.1) *and achieved at* 3.176%.

2.1.1.4 Other Types of Gilts

There still are a couple of other types of gilts but now constituting only a fraction of gilts outstanding in the market, for instance, double-dated gilts and undated gilts. Double-dated gilts have an embedded call option that provides the government with the right to redeem the bond at par value on either final maturity date or some earlier time with three months notice in advance. For the undated gilts, they are the gilts that have not been specified the final redemption date. The redemption date of these bonds is at discretion of the government.

2.1.2 UK Money Market

Three main money market instruments in the UK are presented as follows.

2.1.2.1 BBA LIBOR

LIBOR (London InterBank Offered Rate) is the interest rate at which banks offer to lend unsecured funds to other banks in the London interbank market. It is often used as a key benchmark for short-term interest rates. Since LIBOR rates are usually transacted over-the-counter (OTC) with different market sizes, the British Bankers' Association (BBA) has standardised the rates by launching the fixing procedure and introduced the BBA LIBOR since January 1986. Currently, the BBA LIBOR has been quoted in 10 major currencies with 15 maturities from overnight to 12 months.

Market Convention

The BBA LIBOR is a simple (not compounded) interest rate in which an interest due (for GBP) can be computed by

$$\text{Interest due} = \text{Principal} \times \left(\frac{\text{LIBOR rate}}{100} \right) \times \left(\frac{\text{Actual number of days in interest period}}{365} \right).$$

Note that it is important to carry out the exact number of days in the interest period which is not always, for example, 90 days for a 3-month deposit but could be 89 or 91 days.

Fixings Procedure

The BBA fixes the LIBOR rates by the following procedure.

1. A panel of contributor banks is first selected (and is reviewed at least once a year) based on the reputation, credit quality and activity in London. At present, there are 16 major banks contributing to the GBP panel.

2. On every business day, the BBA gathers the offered rates, for a given standardised notional amount, from all contributor banks for each currency and maturity just prior to 11:00 am (London time).

3. After that, all the rates will be ranked in order and the highest and the lowest quartiles will be excluded. The BBA LIBOR rates of the day will be computed by an arithmetic average on the middle two quartiles.

2.1.2.2 BBA Repo

Repurchase agreement (Repo) is a secured or collateralised loan in which one party agrees to sell securities (e.g. gilts) to the other against a transfer of funds and at the same time the parties agree to repurchase the same or equivalent securities at a specific price in the future. Most Repo rates are transacted on a "general collateral" (GC) basis using the securities such as G-7 government bonds or UK gilts as a collateral. There is also a "special" Repo which is the Repo rate secured by a specific security so that the rates will much rely on the agreed collateral.

The BBA introduced the BBA Repo (only for GBP) as a new benchmark in May 1999. The fixing procedure is generally similar to that for the BBA LIBOR.

2.1.2.3 UK Treasury Bills

UK Treasury Bills are mainly used by the Exchequer as an instrument for cash management operations. They are traded alongside gilts and managed by the DMO. UK Treasury Bills are not as popular as US Treasury Bills which are widely traded by the banks and financial institutions globally.

2.2 The Behaviour of UK Interest rates

Usually, interest rates are considered in three particular ranges of maturities: less than 1 year (short-term), 1 to 10 years (medium-term) and over 10 years (long-term). For the UK market, Strips are frequently used as a benchmark rate for the medium- and long-term. Regarding the short-term, however, investors are likely to manage their investment portfolios through the money market (i.e. Repo and LIBOR rates) rather than the gilt market. We first discuss the three distinct kinds of short-term interest rates in the UK.

2.2.1 Benchmarks for UK Short-Term Interest Rates

Daily UK Strips, BBA Repo and BBA LIBOR rates for 3-month and 6-month maturities from 25 November 2002 to 31 December 2009 are here mainly investigated. Note that for the Strips data, we employ simple linear interpolation to obtain the constant maturities. In Figure 2.1, we can observe that LIBOR rates (unsecured

rates) are typically higher than Strips and Repo rates (secured rates) owing to the realisation of higher risk of default. From the end of 2002 to mid-2007, it can be seen that all the rates slowly climb up from approximately 4% to 6% but afterwards LIBOR rates immediately jumped up to almost 7% in September 2007 (as the US subprime mortgage crisis loomed), while Repo rates and Strips could still remain relatively more stable. From September 2008, all the rates dramatically decline to just above zero since the Bank of England aggressively cut the policy bank rate in order to prevent severe recession from the credit crisis.

Figure 2.2 shows the daily yield changes of all three rates of interest for 3-month and 6-month maturities. It can be noticed from the figure that Strips have clear clusters of low volatility (e.g. around end-2006) and high volatility (e.g. 2008) for both maturities. LIBOR and Repo rates are generally much less volatile and there exist two extreme yield changes on 7 November 2008 and 5 December 2008 when the Bank of England made a big cut on the bank rate from 4.5% to 3.0% and from 3.0% to 2.0% respectively.

Figure 2.3 illustrates the daily spread movements among Strips, Repo and LIBOR rates. Before mid-2007, it can be found that all the spreads move around a narrow range of -0.2% to 0.6%, in which those of LIBOR over Repo rates appear to be most constant at about the level of 0.15% for both maturities. Nevertheless, during the turmoil period (after mid-2007), the spreads are extremely high and wildly fluctuate due to the credit crisis, reflecting an extra premium required for unsecured over secured interest rates.

In summary, each short-term interest rate has its own characteristics but Strips tends to move closer to Repo rates since they both are the secured rates of fund. To select which rates should be used as a proxy for the short-term of a yield curve, it much relies on the model being considered. For a model for long-term investment, either UK Strips or Repo rates would be preferable since they are more stable than LIBOR rates. Particularly, both rates are secured loans which are better representatives for a fund that has a long position in short-term debt. Nonetheless, it is noted that UK Strips tend to have clusters of low and high volatilities clearer than Repo rates and the data is not available for the constant maturities.

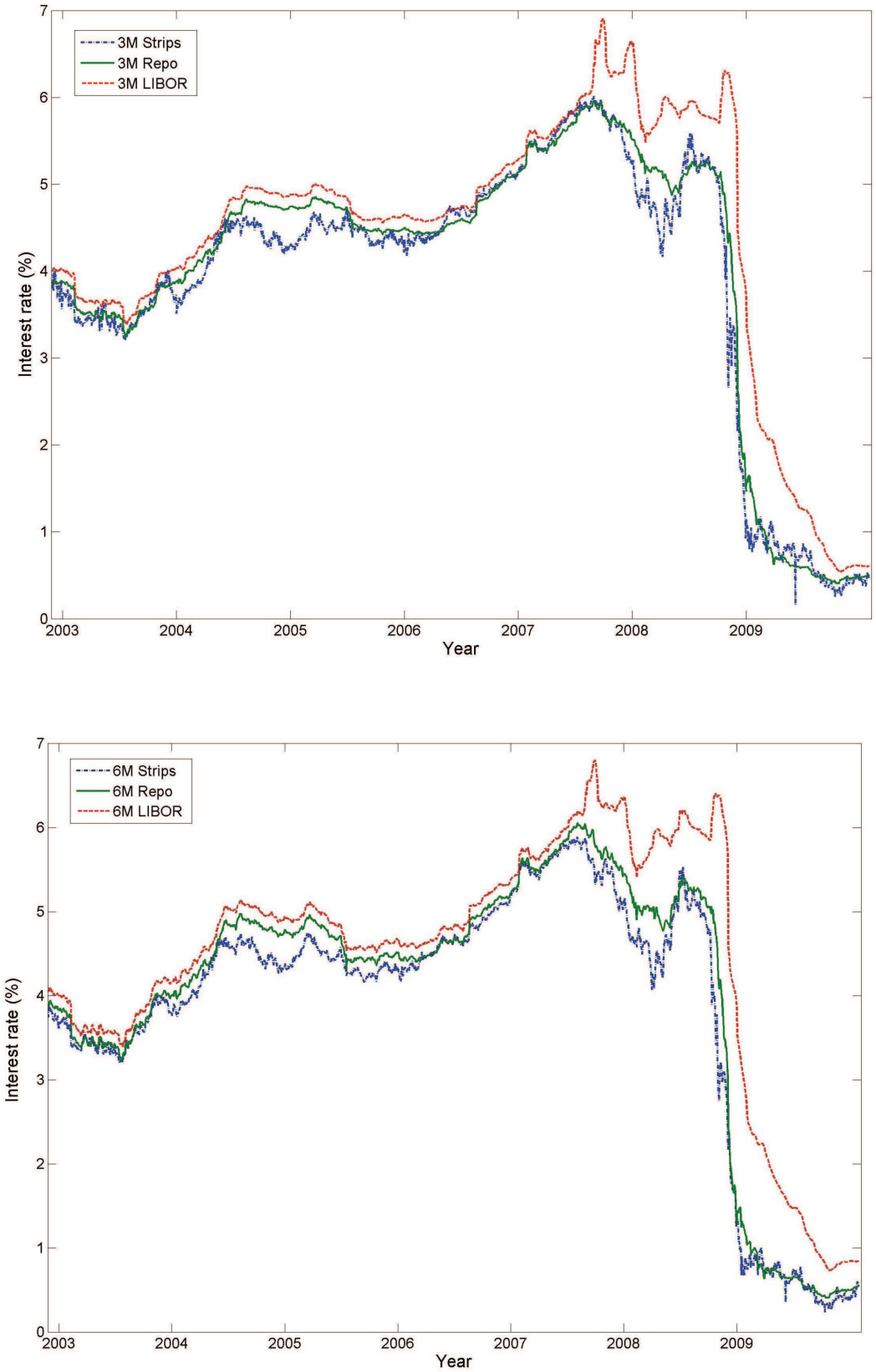

Figure 2.1: Daily UK Strips, BBA Repo and BBA LIBOR (in percentage) for 3-month (top) and 6-month (bottom) maturities from 25 November 2002 to 31 December 2009.

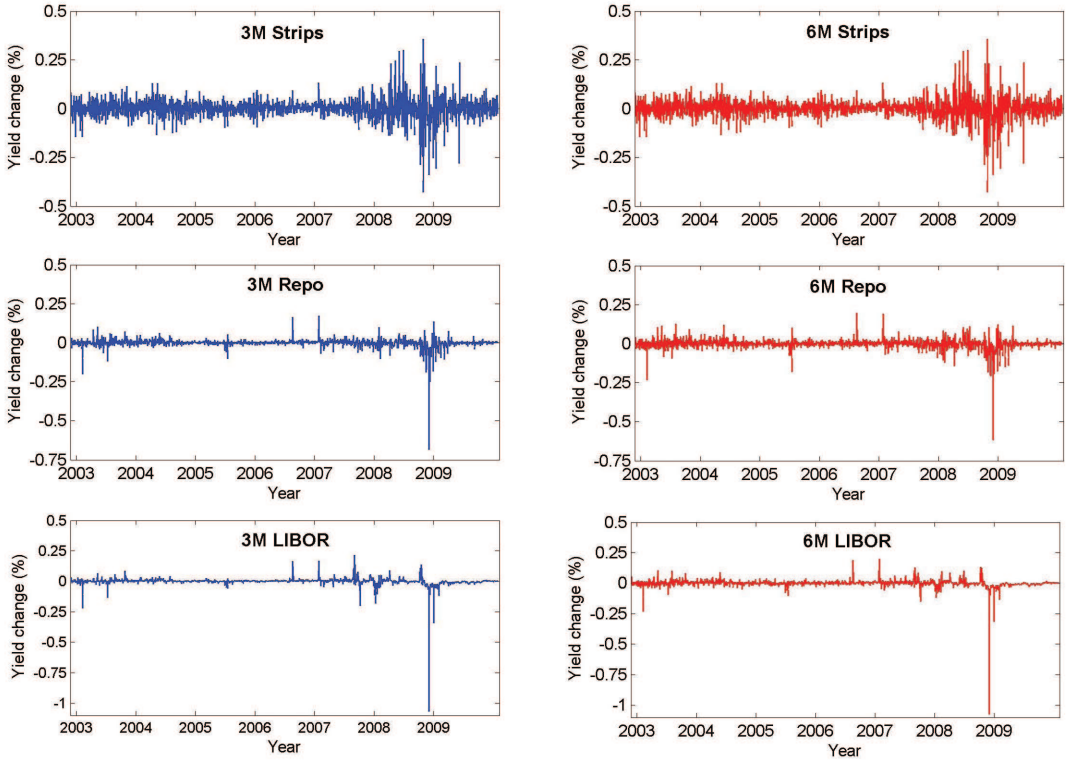

Figure 2.2: Daily yield changes (in percentage) of UK Strips (top), BBA Repo (middle) and BBA LIBOR (bottom) for 3-month (left column) and 6-month (right column) maturities from 25 November 2002 to 31 December 2009.

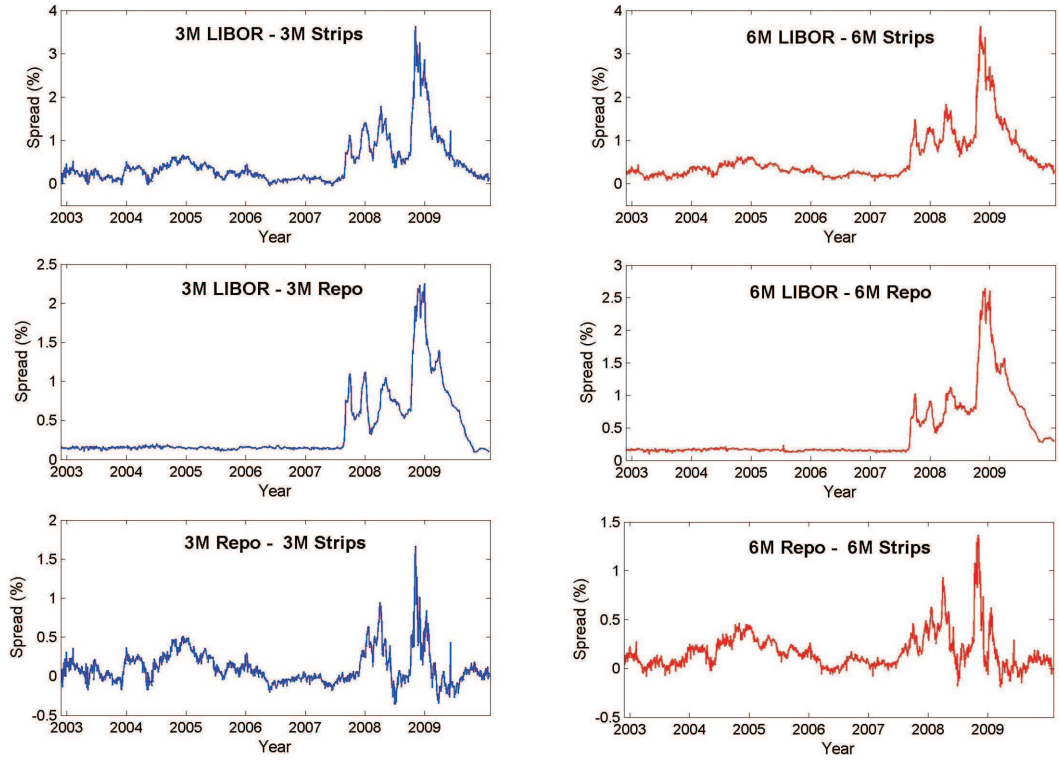

Figure 2.3: Daily spreads (in percentage) among UK Strips, BBA Repo and BBA LIBOR for 3-month (left column) and 6-month (right column) maturities from 25 November 2002 to 31 December 2009.

2.2.2 Short-, Medium- and Long-Term UK Interest Rates

We now analyse the short-, medium- and long-term UK interest rates which are represented by 3-month (3M) Repo, 5-year (5Y) Strips and 30-year (30Y) Strips respectively. As before, the dataset ranges from 25 November 2002 to 31 December 2009 (1,795 observations). Some observations can be made as follows.

- Figure 2.4 shows the daily movement of short-, medium- and long-term yields from 25 November 2002 to 31 December 2009. Unsurprisingly, the long-term are generally more stable than the short- and medium-term interest rates, even during the turmoil period from the second half of 2007. From mid-2008 onwards, the medium-term and especially short-term interest rates sharply fall, while the long-term interest rates are very less volatile. Moreover, it is likely to have an evidence of regime switching for the short-term rates.

- Figure 2.5 illustrates the daily yield changes for short-, medium- and long-term. As can be noticed, the clusters of low and high volatilities of the medium- and long-term interest rates are clearer than those of the short-term. Furthermore, we can see that from mid-2007 the period of high volatility of the medium-term interest rates is evidently longer than the long-term.

- In Figure 2.6, the scatter plots of daily yields among the three rates of interest are provided. From the figure, it can be observed that for the time period being considered, the short- and medium-term interest rates are highly correlated ($\rho = 0.92$), while the long-term interest rates are unlikely to correlate with the others. This imperfect correlation suggests us that one-factor model is unlikely to capture the dynamic of a whole yield curve of the UK interest rates.

- The autocorrelation functions (ACFs) of the daily yield changes and the squared daily yield changes are demonstrated in Figure 2.7. For the short-term interest rates, the ACFs of daily yield changes are significantly different from zero within the 95% confidence level after around the 60th lag, whereas the ACFs of the squared values are much lower. Of the medium- and long-term interest rates, almost all the ACFs of the yield changes remain the values within the confidence level, however the ACFs of the squared values turn to be higher, particularly for the medium-term. To an extent, this is consistent with the clusters of high volatility we noticed before from Figure 2.5.

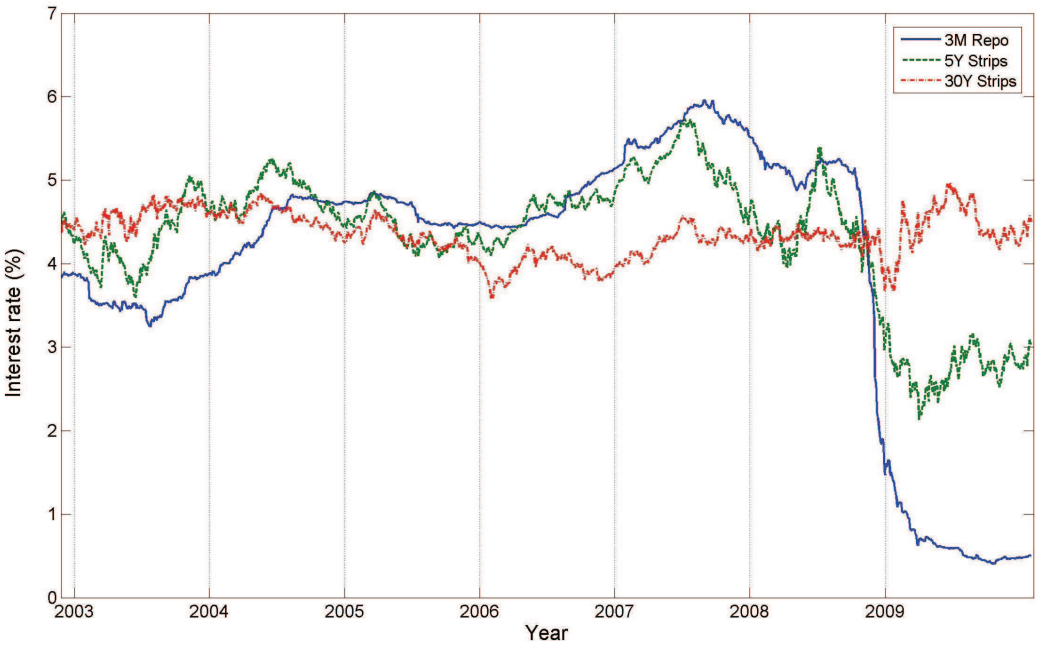

Figure 2.4: Daily movement of UK short-term (3-month Repo), medium-term (5-year Strips) and long-term (30-year Strips) interest rates (in percentage) from 25 November 2002 to 31 December 2009.

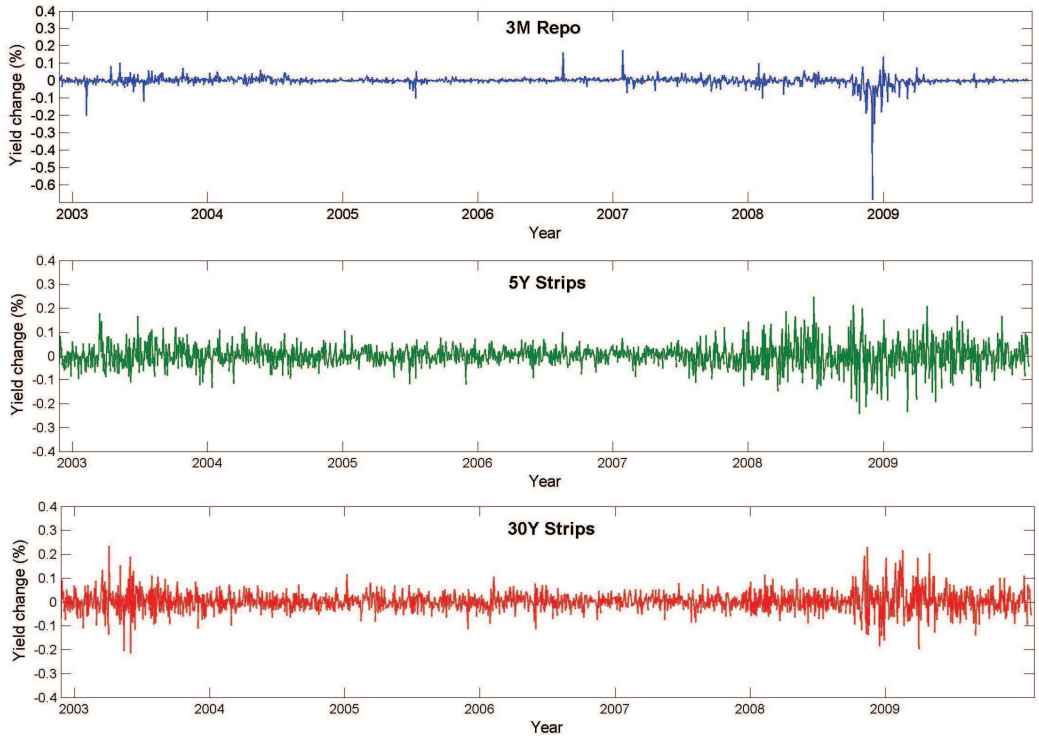

Figure 2.5: Daily yield changes of UK short-term (3-month BBA Repo), medium-term (5-year Strips) and long-term (30-year Strips) interest rates (in percentage) from 25 November 2002 to 31 December 2009.

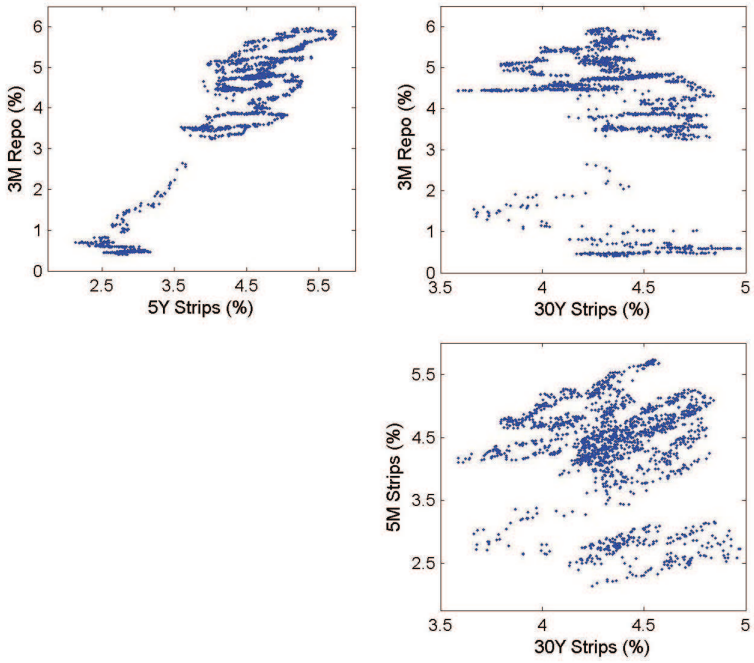

Figure 2.6: Scatter plots of daily UK interest rates (in percentage) from 25 November 2002 to 31 December 2009. Top-left: short-term versus medium-term. Top-right: short-term versus long-term. Bottom-right: medium-term versus long-term.

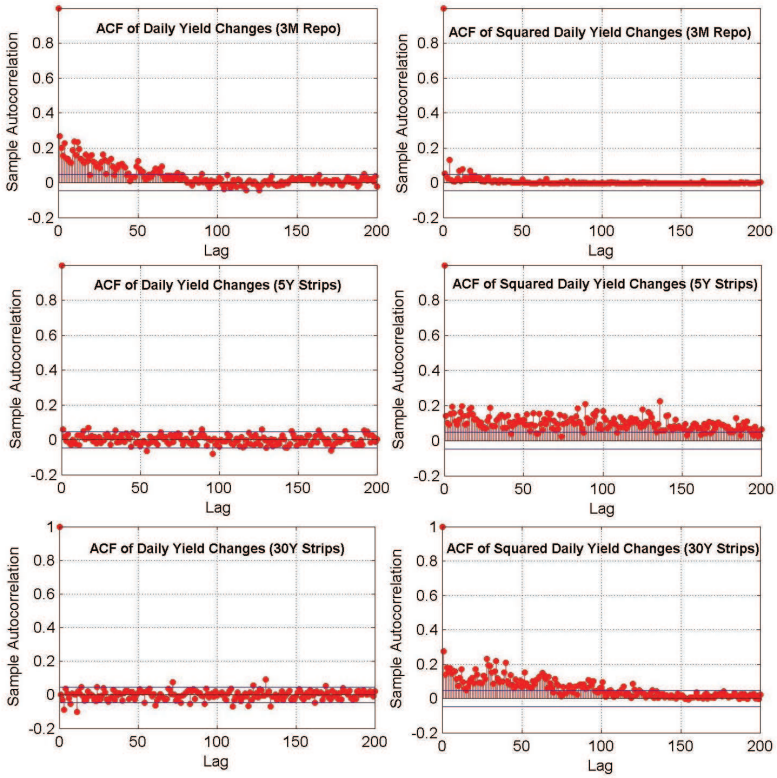

Figure 2.7: Autocorrelation functions of daily yield changes and squared daily yield changes of UK short-term (3-month BBA Repo), medium-term (5-year Strips) and long-term (30-year Strips) interest rates (in percentage) from 25 November 2002 to 31 December 2009.

2.3 Fitting the Nelson-Siegel Model to UK Interest Rates

In this section, we estimate the Nelson-Siegel model with monthly UK Strips using the ordinary least squares (OLS) method. The main purpose is to observe the three latent driving factors (level, slope and curvature) as interpreted by Diebold and Li (2006). To begin with, we recall that the Nelson-Siegel spot rate at time t maturing at T is given by

$$(2.2)\quad R(t,T) = L(t) + \left(\frac{1-e^{-\lambda(t)(T-t)}}{\lambda(t)(T-t)}\right)S(t) + \left(\frac{1-e^{-\lambda(t)(T-t)}}{\lambda(t)(T-t)} - e^{-\lambda(t)(T-t)}\right)C(t),$$

where $L(t), S(t), C(t)$ and $\lambda(t)$ are the time-varying model parameters.

For simplicity and tractability of the estimation, we follow Diebold and Li by fixing $\lambda(t)$ (denoted as λ). Then, the latent factors $L(t), S(t)$ and $C(t)$ (i.e. long-, short- and medium-term factors) will be estimated for each time t. Regarding the suitable value of λ, we consider two possible values: λ_1 (the value achieved by minimising the sum of squared residuals for a whole yield surface of the data) and $\lambda_2 = 0.0299$ (the value at which the loading on $C(t)$ is at maximum, given $T - t = 5$ years).

Additionally, the three driving factors may also be defined from the data. According to (2.2), It can be proved that when $T - t \to \infty, R(t,\infty) = L(t)$ and hence the level factor might be equivalent to the long-term yield. Furthermore, for the short-term or slope factor, theoretically it can be defined as $R(t,\infty) - R(t,t) = -S(t)$. For the medium-term or curvature factor, it may be approximated as the twice the medium-term yield subtracted by the sum of the short- and long-term yields.

2.3.1 Results

Monthly UK Strips yields from November 2002 to June 2008 (68 months) are considered for estimating the Nelson-Siegel model using the OLS method. The dataset is of the last business day of each month and is pooled into fixed 20 maturities (0.25, 0.5, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 12.5, 15.0, 17.5, 20.0, 22.5, 25.0, 27.5, 30.0 years) by simple linear interpolation. As can be seen from Figure 2.8, we can initially notice that in the first half of the time period considered, the yield curves

are generally upward-sloping. However, they subsequently turn to be flatter and then inverted. In addition, it is noted that the short-end is more volatile than the long-end of the yield curves. From this data, we obtain $\lambda_1 = 0.0135$ where the sum of squared residuals is at the minimum.

Figure 2.9 shows the estimated three latent driving factors $L(t), S(t)$ and $C(t)$ with $\lambda_1 = 0.0135$ (left column) and $\lambda_2 = 0.0299$ (right column) comparing to those defined from the data (30Y Strips, 3M Strips minus 30Y Strips and the twice the 5Y Strips minus the sum of 3M Strips and 30Y Strips respectively). As can be seen, λ_2 provides the estimated driving factors much closer to those defined from the data than λ_1.

Regarding the loadings on the three driving factors, we can observe from Figure 2.10 that in both cases the loading on $L(t)$ is constant at 1 which can be regarded as a long-term factor. For the loading on $S(t)$, $(1-e^{-\lambda(T-t)})/\lambda(T-t)$, it roughly begins at 1 but decays monotonically to approximately 0.2 and 0.1 at 30-year maturity for λ_1 and λ_2 respectively (λ_2 results in the faster rate of decay than λ_1). This may be viewed as a short-term factor. Finally, the loading on $C(t)$, $((1-e^{-\lambda(T-t)})/\lambda(T-t)) - e^{-\lambda(T-t)}$, starts at about 0, increases to maximise at 10-year maturity for λ_1 and 5-year maturity for λ_2, and then decays to the values very close to the loadings on $S(t)$ at 30-year maturity. Hence, it may be interpreted as a medium-term factor.

We conclude here that all the above results assert Diebold and Li's interpretation that the three parameter models in the Nelson-Siegel model correspond to level, slope and curvature factors. In addition, the correlations between the estimated factors $\rho(L(t), S(t))$, $\rho(L(t), C(t))$ and $\rho(S(t), C(t))$ are found to be -0.81, -0.28 and -0.07 for λ_1, and -0.80, -0.35 and 0.33 for λ_2.

In Figure 2.11, the yield residual surfaces given λ_1 (top) and λ_2 (bottom) are illustrated. The sum of squared residuals for a whole surface is achieved at 6.42 and 8.10 respectively. As can be seen, the surface with λ_1 generally appears to be flatter. However, from 2008 onwards the residuals in both cases are substantial high for almost all maturities. Moreover, it is likely to have some particular pattern for the residuals with λ_2 as we can notice that all residuals are high at the short end and between the maturities of around 10 to 20 years over a whole time period.

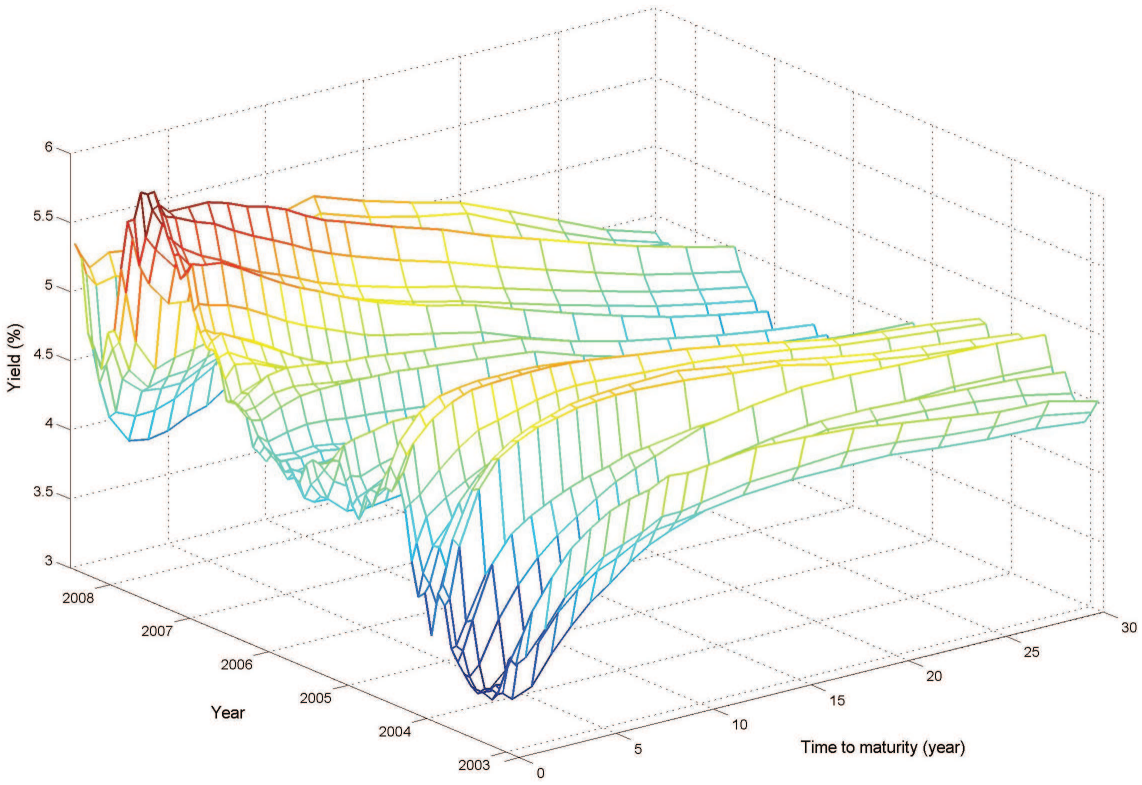

Figure 2.8: UK Strips yield surface from November 2002 to June 2008 (68 months).

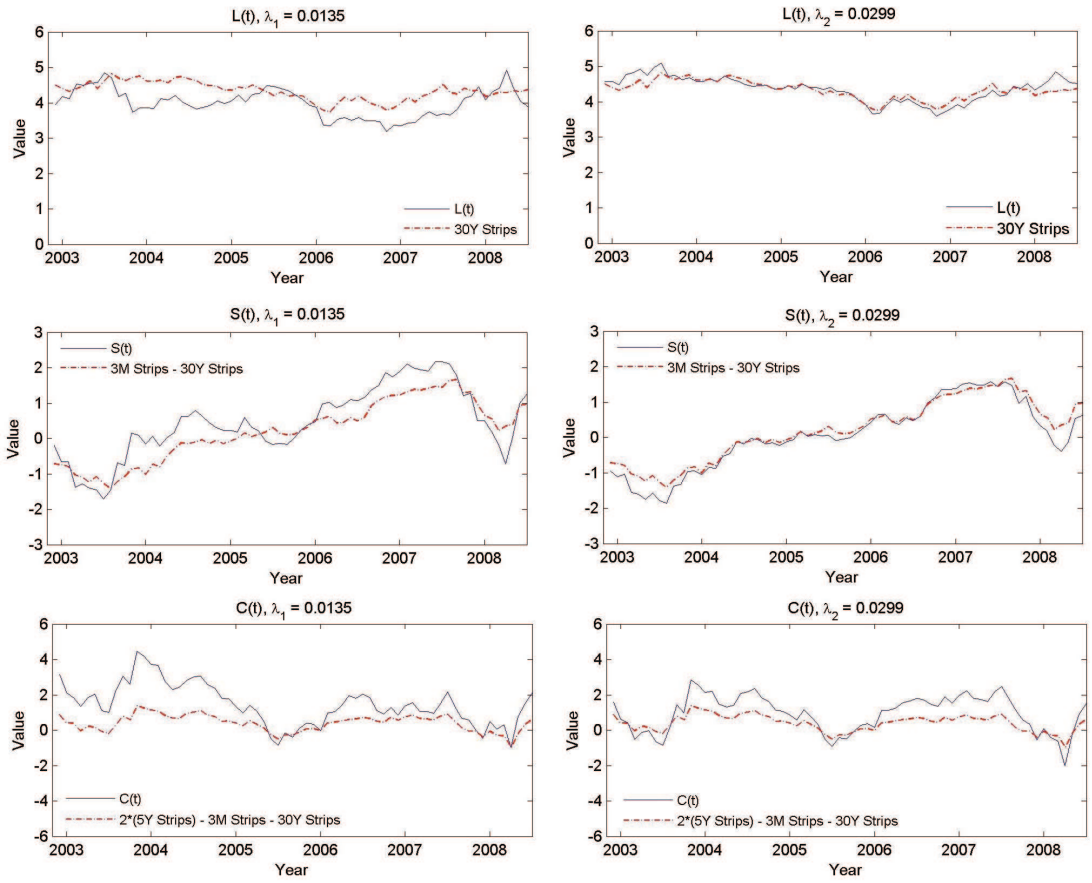

Figure 2.9: Comparison of the latent driving factors defined from the data (dot-dashed) and the estimates using the OLS method (solid): level $L(t)$ (top), slope $S(t)$ (middle) and curvature $C(t)$ (bottom) of the Nelson-Siegel model fitted with monthly UK Strips from November 2002 to June 2008 with $\lambda_1 = 0.0135$ (left column) and $\lambda_2 = 0.0299$ (right column).

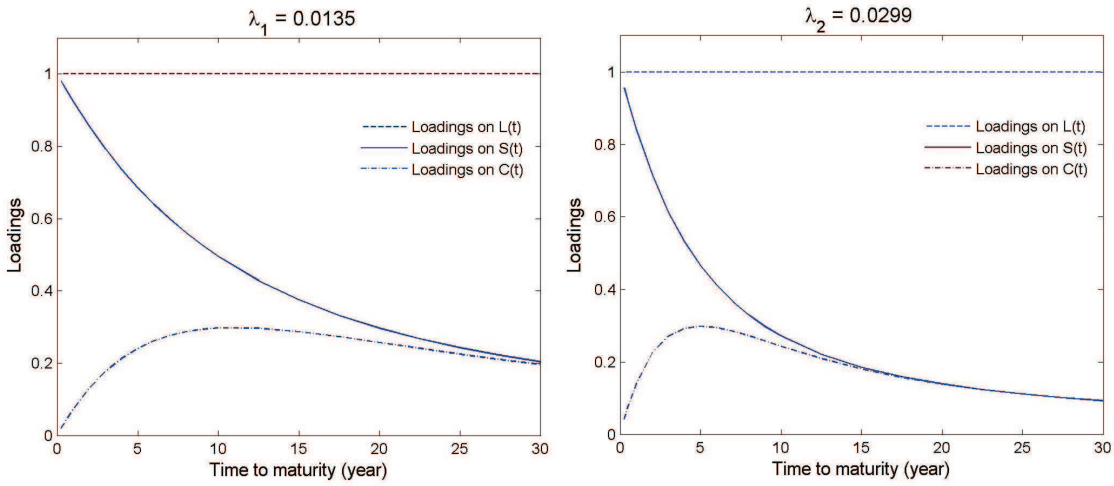

Figure 2.10: Loadings on $L(t)$, $S(t)$ and $C(t)$ by $\lambda_1 = 0.0135$ (left column) and $\lambda_2 = 0.0299$.

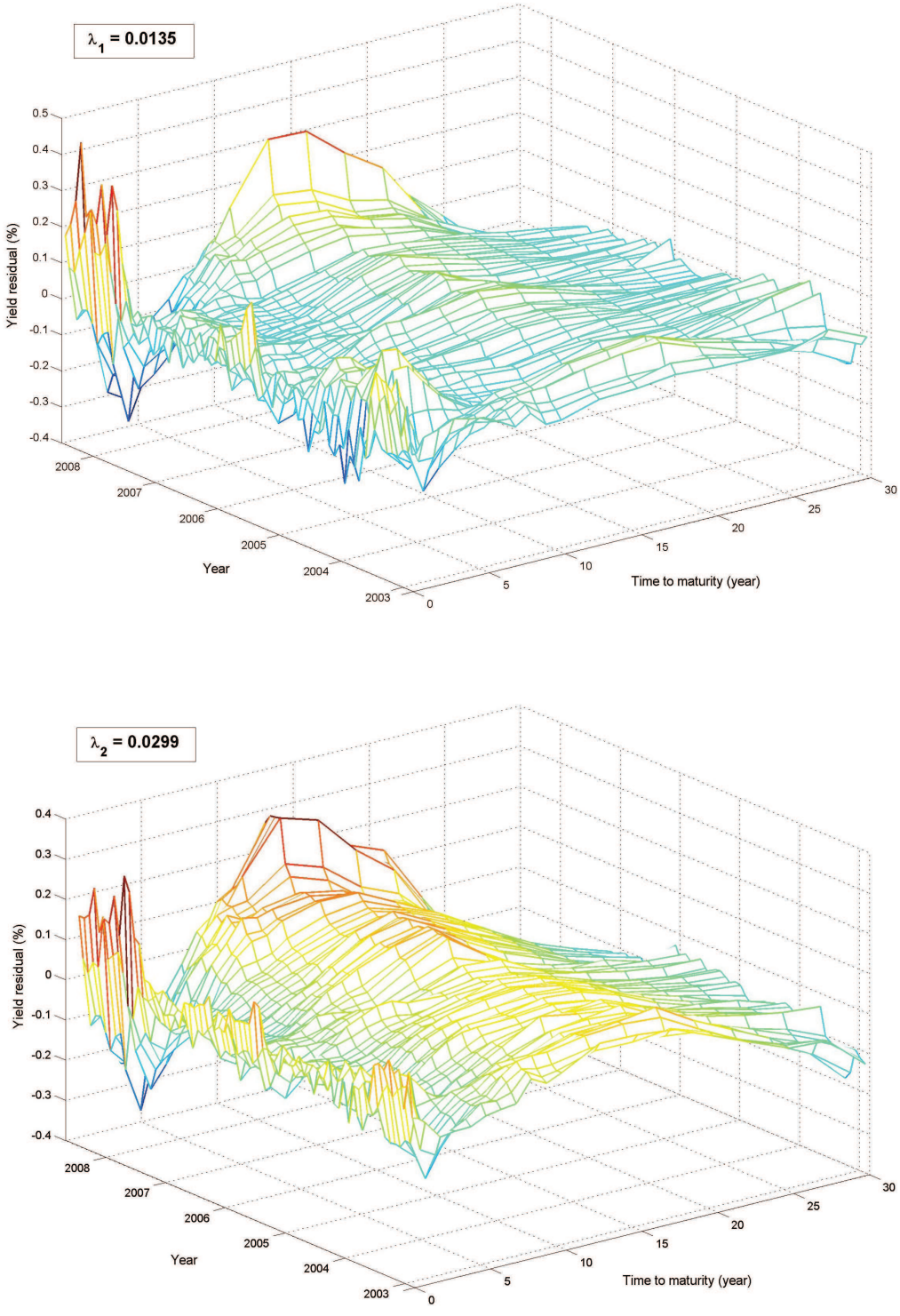

Figure 2.11: Yield residual surfaces of the Nelson-Siegel model fitted with monthly UK Strips from November 2002 to June 2008 with $\lambda_1 = 0.0135$ (top) and $\lambda_2 = 0.0299$ (bottom).

2.4 Principal Components Analysis

Principal components analysis (PCA) is a mathematical technique used for reducing the dimensions of large data into a smaller number of main principal components which are uncorrelated by the orthogonality property. In term structure modelling, the PCA methodology can be employed to construct an empirical volatility structure and the procedure is as below.

- Take yield changes of the interest rate data.
- Compute the covariance matrix of the yield changes.
- Determine the eigenvectors and eigenvalues of the covariance matrix.

Each eigenvalue can be interpreted as the variance of the corresponding eigenvector. As a property of the orthogonality, the sum of all eigenvalues is equal to the total variance. Typically, the first three main components (i.e. three largest eigenvalues) will be considered, in which from the empirical evidence, the accumulated variance is usually around 90% to 95% of the total variance.

2.4.1 PCA Results on Daily UK Interest Rates

We first note that the PCA approach does not allow jumps in the data (James and Webber, 2000). Therefore, we initially consider the daily UK interest rates from 25 November 2002 to 30 May 2008 since we already noticed that from mid-2008 onwards, short-term UK interest rates dramatically decline. In addition, we also use Repo rates in place of UK Strips to represent the short-term interest rates. Specifically, we are conducting PCA analysis to daily UK interest rates of 10 constant maturities: BBA Repo (3M, 6M, 1Y) and UK Strips (2Y, 3Y, 5Y, 7Y, 10Y, 20Y, 30Y).

Table 2.1 shows all ten principal components or the eigenvalues and eigenvectors obtained from the PCA on the daily UK interest rates. The correlation matrix of the yield changes is provided in Table 2.2. As can be seen from the results, the first three components account for almost all the variance of the data, i.e. the accumulated variance of the first three components is equal to 96.54%. Furthermore, from this result we may infer that three (or at least two) randomnesses or factors should be included in a model for UK interest rates.

The first (most impact) three main components are also plotted in Figure 2.12. From the figure, it can be understood that the first component moves a whole yield curve almost equally likely, while the second component generally drives the short- and long-term interest rates move in opposite direction. For the third component, it has largest positive impact on medium-term interest rates.

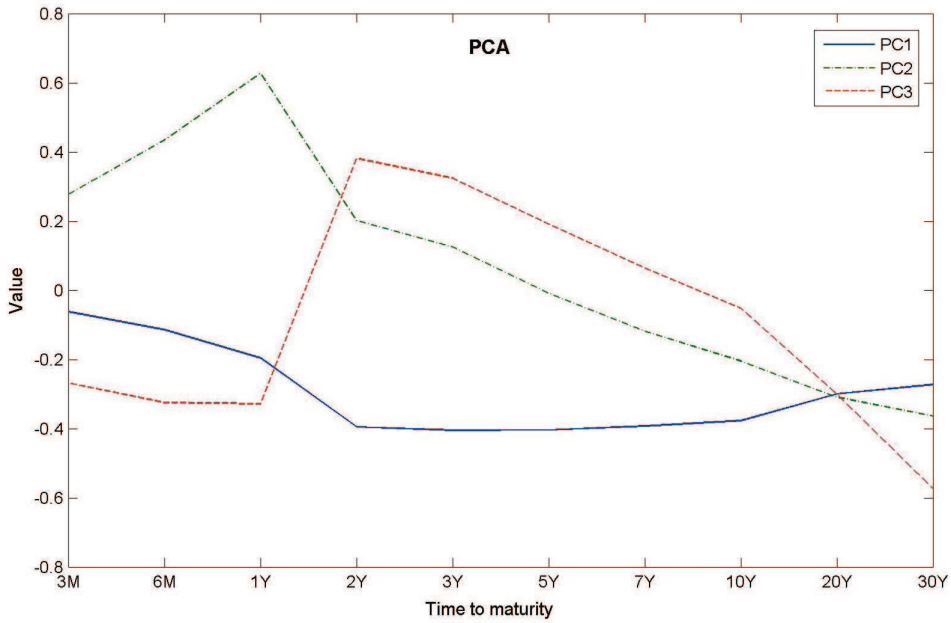

Figure 2.12: Three main principal components from daily UK interest rates data from 25 November 2002 to 30 May 2008 of the 10 maturities.

Maturity	Principal Components									
	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
3M	-0.062	0.278	-0.268	0.097	-0.746	0.068	-0.226	-0.473	0.016	-0.016
6M	-0.114	0.434	-0.325	0.076	-0.258	-0.030	0.348	0.706	0.007	-0.003
1Y	-0.196	0.627	-0.328	0.049	0.597	-0.014	-0.145	-0.284	-0.015	0.008
2Y	-0.395	0.201	0.381	-0.455	-0.110	-0.458	0.410	-0.225	0.104	0.014
3Y	-0.405	0.125	0.324	-0.210	-0.076	0.181	-0.531	0.277	-0.502	0.147
5Y	-0.404	-0.008	0.191	0.103	0.011	0.317	-0.164	0.106	0.589	-0.551
7Y	-0.393	-0.119	0.065	0.361	0.015	0.253	0.179	-0.070	0.254	0.729
10Y	-0.377	-0.205	-0.053	0.434	0.027	0.038	0.375	-0.175	-0.557	-0.376
20Y	-0.300	-0.309	-0.301	0.195	0.006	-0.708	-0.384	0.143	0.127	0.028
30Y	-0.273	-0.364	-0.575	-0.603	0.038	0.292	0.096	-0.041	-0.009	0.017
Eigenvalue ($\times 10^{-2}$)	0.995	0.194	0.068	0.019	0.012	0.006	0.003	0.003	0.002	0.001
(%)	76.46	14.89	5.19	1.42	0.94	0.43	0.24	0.21	0.12	0.09
Accumulated (%)	76.46	91.35	96.54	97.97	98.91	99.34	99.58	99.79	99.91	100.00

Table 2.1: Eigenvectors and values from PCA analysis on daily UK interest rates data from 25 November 2002 to 30 May 2008 of the 10 maturities.

Maturity	3M	6M	1Y	2Y	3Y	5Y	7Y	10Y	20Y	30Y
3M	1.000	0.872	0.743	0.376	0.351	0.292	0.241	0.195	0.120	0.098
6M		1.000	0.921	0.515	0.480	0.412	0.343	0.284	0.176	0.135
1Y			1.000	0.610	0.578	0.508	0.429	0.360	0.227	0.167
2Y				1.000	0.975	0.936	0.875	0.816	0.672	0.544
3Y					1.000	0.972	0.925	0.872	0.735	0.603
5Y						1.000	0.978	0.945	0.831	0.696
7Y							1.000	0.982	0.895	0.757
10Y								1.000	0.937	0.808
20Y									1.000	0.902
30Y										1.000

Table 2.2: Correlation matrix of daily UK yield changes from 25 November 2002 to 30 May 2008 of the 10 maturities.

2.4.2 A Use of the Principal Components

To use the achieved PCA results, we set up the yield equation

$$X(t+1) = X(t) + MSZ(t+1), \tag{2.3}$$

where $X(t+1) = (X^{3m}, X^{6m}, \ldots, X^{30y})'$ is the observed yields of the 10 maturities at time $t+1$, $Z(t+1) = (Z_1(t+1), \ldots, Z_{10}(t+1))'$ is the i.i.d. unit normal random variables at time $t+1$, M and S are the principal component and volatility matrices obtained from the PCA analysis in Table 2.1 such that

$$M = \begin{pmatrix} pc_1^{3m} & pc_2^{3m} & \cdots & pc_{10}^{3m} \\ pc_1^{6m} & pc_2^{6m} & \cdots & pc_{10}^{6m} \\ \vdots & \vdots & \ddots & \vdots \\ pc_1^{30y} & pc_2^{30y} & \cdots & pc_{10}^{30y} \end{pmatrix}, S = \begin{pmatrix} \sqrt{\lambda}_1 & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda}_{10} \end{pmatrix},$$

where λ_i, for $i = 1, 2, ...10$, are the eigenvalues or the variances of principal component i and $M^T M = I$.

Specifically, we have

$$\begin{bmatrix} \delta^{3m}(t+1) \\ \delta^{6m}(t+1) \\ \vdots \\ \delta^{30y}(t+1) \end{bmatrix} = \begin{bmatrix} pc_1^{3m}\sqrt{\lambda}_1 Z_1(t+1) + pc_2^{3m}\sqrt{\lambda}_2 Z_2(t+1) + \cdots + pc_{10}^{3m}\sqrt{\lambda}_{10} Z_{10}(t+1) \\ pc_1^{6m}\sqrt{\lambda}_1 Z_1(t+1) + pc_2^{6m}\sqrt{\lambda}_2 Z_2(t+1) + \cdots + pc_{10}^{6m}\sqrt{\lambda}_{10} Z_{10}(t+1) \\ \vdots \\ pc_1^{30y}\sqrt{\lambda}_1 Z_1(t+1) + pc_2^{30y}\sqrt{\lambda}_2 Z_2(t+1) + \cdots + pc_{10}^{30y}\sqrt{\lambda}_{10} Z_{10}(t+1) \end{bmatrix},$$

where $\delta(t+1) = X(t+1) - X(t)$.

From (2.3), the standardised residuals of the principal components at time $t+1$, $\hat{Z}(t+1) = (\hat{Z}_1(t+1), \ldots, \hat{Z}_{10}(t+1))'$ can be defined by

$$\hat{Z}(t+1) = S^{-1}M^{-1}\delta(t+1). \tag{2.4}$$

2.4.3 Residual Analysis

Figures 2.13, 2.14 and 2.15 illustrate the standardised residuals of the first three principal components $\hat{Z}_1(t)$, $\hat{Z}_2(t)$ and $\hat{Z}_3(t)$ respectively. From the figures (top left),

we can see that there exist clusters of high and low volatilities on the standardised residuals of all three components. Moreover, a number of extreme values appear particularly on the series of $\hat{Z}_2(t)$ and $\hat{Z}_3(t)$. Also, it can be noticed from the QQ-plots (top right) that all $\hat{Z}_1(t), \hat{Z}_2(t)$ and $\hat{Z}_3(t)$ are clearly not normally distributed and have the fatter tails. Regarding the autocorrelation functions (bottom left and right), those of $\hat{Z}_1(t)^2$ are moderately high whereas those of $\hat{Z}_1(t), \hat{Z}_2(t), \hat{Z}_3(t), \hat{Z}_2(t)^2$ and $\hat{Z}_3(t)^2$ remain rather low although the values at some lags are out of the 95% confidence level. Finally, it is noted that using the empirical volatility structure from the PCA analysis according to (2.3) is evidently inadequate to capture features of the UK interest rates since the standardised residuals of the main components still do not reasonably look i.i.d. (i.e. there is evidence for clusters over time). An extension of PCA combines this with a GARCH model (Engle, 1982). Such methodology is known as the principal component GARCH or orthogonal GARCH (O-GARCH). More details can be found in Alexander (2002).

In summary, Chapter 2 introduced and analysed UK interest rate data which will be used throughout the thesis. Fitting the Nelson-Siegel model and the principal components analysis motivated and justified the use of a multifactor model for interest rates.

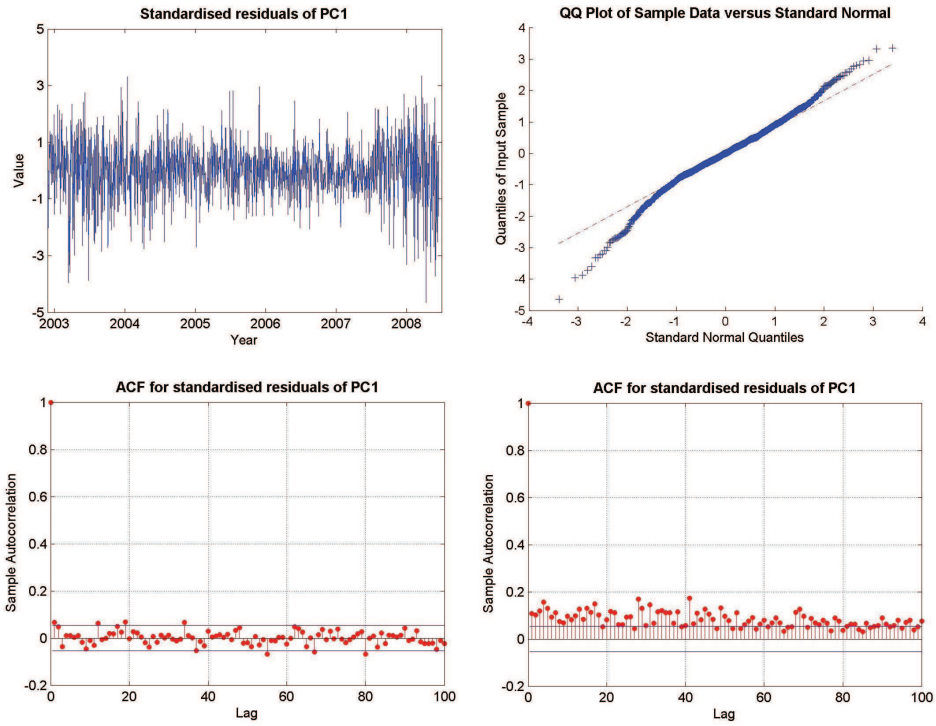

Figure 2.13: Standardised residuals, $\hat{Z}_1(t)$ (top left), normal QQ-plot (top right) and the autocorrelation functions for the standardised (bottom left) and the squared standardised residuals of the first principal component.

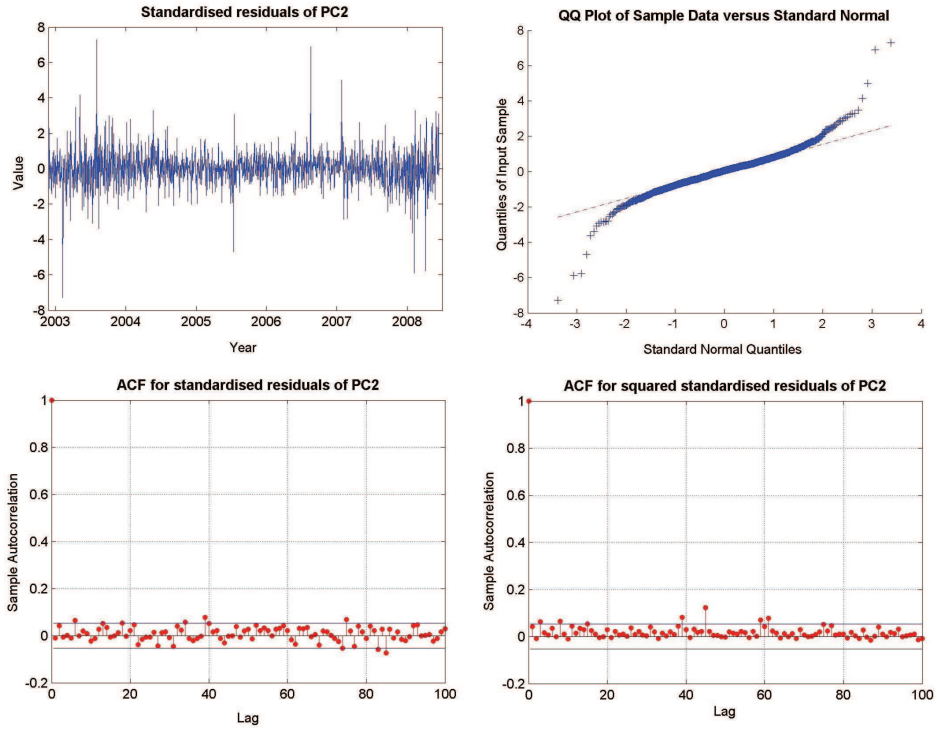

Figure 2.14: Standardised residuals, $\hat{Z}_2(t)$ (top left), normal QQ-plot (top right) and the autocorrelation functions for the standardised (bottom left) and the squared standardised residuals of the second principal component.

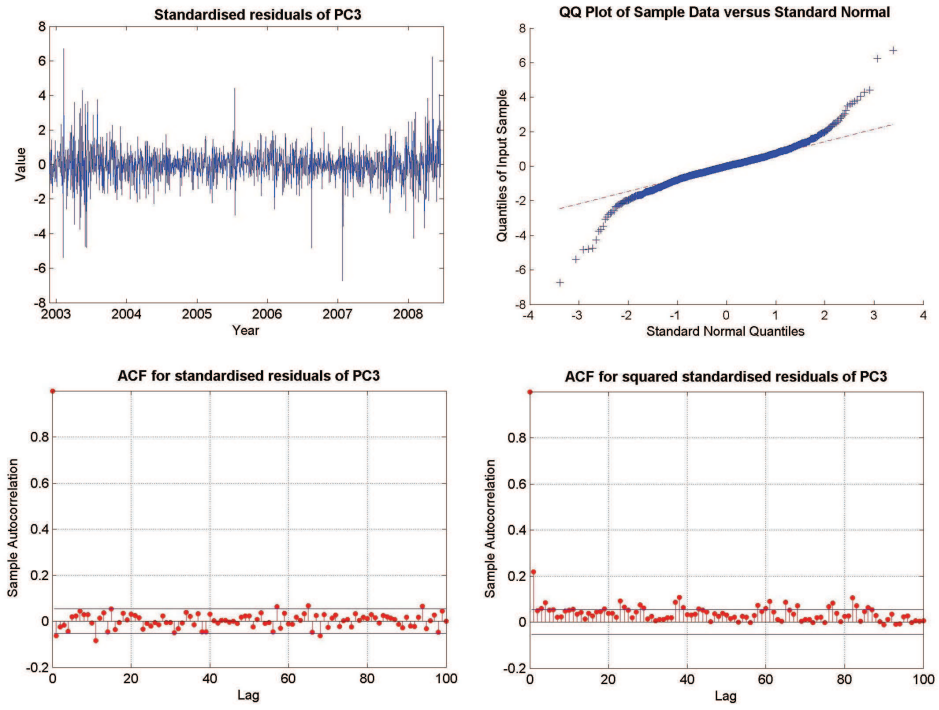

Figure 2.15: Standardised residuals, $\hat{Z}_3(t)$ (top left), normal QQ-plot (top right) and the autocorrelation functions for the standardised (bottom left) and the squared standardised residuals of the third principal component.

Chapter 3

Estimation Techniques in Term Structure Modelling

This chapter presents some modern estimation techniques having been used in term structure modelling. We first provide an overview of the methods of our interest which are maximum likelihood (ML), generalised method of moments (GMM), efficient method of moments (EMM) and Markov chain Monte Carlo (MCMC). Each method is consequently described but MCMC will be discussed in great detail in the next chapter. To implement EMM, the semi-nonparametric (SNP) conditional density model that serves as a general auxiliary model is also discussed. In the end, we use the GMM and EMM methodologies to estimate a one-factor CKLS model with daily 3-month UK Repo rates mainly for illustration purpose.

3.1 Introduction

It has been known that the movement of interest rates is usually highly persistent and close to a non-stationary process with a unit root (Ball and Torous, 1996). In addition, the simple estimation method such as the ordinary least squares (OLS) is far from appropriate for calibrating sophisticated term structure models since we typically wish to have the residuals be i.i.d. rather than to minimise them. Several advanced statistical methods were therefore developed in order to cope with this complexity.

At an early stage, likelihood-based estimation played a key role in statistical inference and modelling. It was widely used in many applications, including the estimation

of term structure models. For instance, Pearson and Sun (1994) and Nowman (1997) use the maximum likelihood (ML) method to estimate a two-factor CIR model and a one-factor CKLS model respectively. After a while, much of the attention moved to the general method of moments (GMM) when it was first formalised by Hansen (1982). The GMM method generalises the standard method of moments (MM) in which the number of moment functions can be greater than the number of parameters being estimated. More precisely, the moment functions are initially defined and then we solve an optimisation problem of equating the sample average of the moment functions to zero. In case that the number of moment functions is just equal to the number of parameters, the exact solution thus can be achieved. Examples of applications to term structure modelling are those by Longstaff and Schwartz (1992) and Chan et al. (1992), where they apply the GMM methodology to estimate the CIR and CKLS models respectively.

There are a number of advantages claimed for the GMM over ML method (Jagannathan and Wang, 2002). First, GMM is more straightforward, generic and convenient to use. It relies solely on the moment conditions and makes no use of knowledge of the distribution which is required for ML. Furthermore, GMM provides a model misspecification test and diagnostics inherently from the framework setting when there are more moment functions than unknown parameters. In contrast, the ML method is typically required to derive the test on a model basis. Nevertheless, in spite of these advantages, one should bear in mind that when the distributional assumptions are appropriately defined, ML tends to provide the most efficient estimates of model parameters, while GMM may not. Moreover, GMM normally requires the time series to be observed for long enough for its moments to have converged. Most importantly, since the GMM method much depends on the moment functions which have a considerable number of choices, inappropriate selection of the moment conditions may lead to the biased and inefficient estimation. Therefore, to decide which methods to use, one must consider characteristics and assumptions of the model being calibrated carefully.

For the models incorporated with unobserved or latent variables, the GMM method may not be applicable if the moment functions cannot be numerically evaluated. Additionally, in a presence of the latent variables, the complete likelihood

function may also be hard to obtain. In that case, the ML method is unlikely to be feasible. Under these circumstances, one may use the simulated method of moments (SMM) for estimation. The technical properties of SMM methodology can be found in Duffie and Singleton (1993). The efficient method of moments (EMM), described by Gallant and Tauchen (1996), is one kind of the SMM method. Since introduced, it has appeared in several well-known literatures of estimating term structure models, for example, stochastic volatility models (Andersen and Lund, 1996), affine term structure models (Dai and Singleton, 2000) and quadratic term structure models (Ahn et al., 2002). Despite its popularity, one prevailing shortcoming of EMM is computationally expensive comparing to GMM and ML.

In recent years, Bayesian estimation has also increasingly been paid attention due to the development of the Markov Chain Monte Carlo (MCMC) simulation. In the past, the Bayesian approach was less preferable in many cases because of the difficulty of implementation, particularly for the high-dimensional problems. The existence of the MCMC methodology allows us to tackle such problems in more flexible and feasible ways. One distinct advantage of MCMC is that we can obtain information about parameter uncertainty directly from the simulation output. Specifically, MCMC avoids relying on an asymptotic approximation as do GMM and EMM. Nevertheless, implementing MCMC is extremely time-consuming even comparing to EMM. Despite its huge development, MCMC has not appeared much in the literature of term structure modelling. Nonetheless, it will be our main methodology for estimating term structure models in subsequent chapters.

3.2 Maximum Likelihood Estimation

The classical maximum likelihood (ML) estimation assumes that the crucial information of observed data is summarised in an unknown parameter vector θ of the likelihood function. This method therefore aims at achieving θ that maximises the likelihood function, i.e. contains most information of the data.

Definition 3.1. **[Likelihood Function]** Suppose that observed data $y = (y_1, \ldots, y_T)$ are i.i.d. and generated from a model with a fixed and unknown parameter vector θ. Then, the likelihood function $L(\theta|y)$ is defined as the joint probability density

function of y conditioned on the parameter vector θ, $f(y|\theta)$, i.e.

$$L(\theta|y) = f(y_1, \ldots, y_T|\theta) = \prod_{t=1}^{T} f(y_t|\theta) \tag{3.1}$$

and the log-likelihood function is

$$\log L(\theta|y) = \log f(y_1, \ldots, y_T|\theta) = \sum_{t=1}^{T} \log f(y_t|\theta). \tag{3.2}$$

In addition, a quasi maximum likelihood estimator is frequently referred to as an ML estimator of a misspecified model (e.g. we assume the correct dynamic form of a model is fitted but that the innovations are erroneously assumed to be Gaussian) and therefore the estimates obtained are QMLEs (See McNeil et al., 2005).

3.3 Generalised Method of Moments

Referring to Zivot and Wang (2005), the underlying concept of generalised method of moments (GMM) is the orthogonality condition of the moment functions, which may also be associated with some instrumental variables. Without loss of generality, we consider a linear model

$$y_t = x_t'\theta + \varepsilon_t, \quad t = 1, \ldots, T, \tag{3.3}$$

where y_t are observed data, x_t is an $L \times 1$ vector of explanatory variables, θ is the model parameter vector and ε_t is an error term which may be correlated to x_t.

Next, we define ψ_t as a $K \times 1$ vector of instrumental variables and the moment function $f_t(y_t, x_t, \psi_t, \theta) = \psi_t\varepsilon_t = \psi_t(y_t - x_t'\theta)$. Then, the unique GMM estimator of θ is $\hat{\theta}$ that satisfies

- the moment condition:

$$\mathbb{E}[f_t(y_t, x_t, \psi_t, \theta)] = \mathbb{E}[\psi_t\varepsilon_t] = \mathbb{E}[\psi_t(y_t - x_t'\theta)] = 0, \tag{3.4}$$

- the rank condition:

$$rank(\mathbb{E}[\psi_t x_t']) = L, \quad \text{and} \tag{3.5}$$

- the order condition:

$$K \geq L. \tag{3.6}$$

In case that $K = L$, it is commonly known as the "just-identified case" and hence the GMM estimator $\hat{\theta}$ can be achieved by matching the theoretical to the sample moments, i.e.

$$\frac{1}{T}\sum_{t=1}^{T} f_t(y_t, x_t, \psi_t, \hat{\theta}) = 0. \tag{3.7}$$

If $K > L$, "over-identified case", the GMM estimator will be defined by

$$\hat{\theta}(W) = \arg\min_{\theta} \; T \cdot f_T(\theta)' W f_T(\theta), \tag{3.8}$$

where

$$f_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} f_t(y_t, x_t, \psi_t, \theta)$$

and W is some positive definite $K \times K$ symmetric weight matrix.

Theorem 3.2. ***[Asymptotic Distribution of the GMM Estimator]*** *Under some regularity conditions and the model setting in* (3.3)*, the GMM estimator* $\hat{\theta}(W)$ *of* θ *has the following asymptotic properties:*

1. $\hat{\theta}(W) = \theta$ *as* $T \to \infty$.
2. $\sqrt{T}\ (\hat{\theta}(W) - \theta) \xrightarrow{d} N(0, avar(\hat{\theta}(W)))$*, where* $avar(\hat{\theta}(W))$ *is the asymptotic variance equal to* $(\Gamma' W \Gamma)^{-1}\Gamma' W S W \Gamma (\Gamma' W \Gamma)^{-1}$*, where* $\Gamma = \mathbb{E}[\psi_t x_t]$ *and* $S = \mathbb{E}[f_t f_t'] = \mathbb{E}[\psi_t \psi_t' \varepsilon_t^2]$.

Proposition 3.3. ***[The Optimal Weight Matrix]*** *The weight matrix is optimal if and only if* $W = S^{-1}$ *where the asymptotic variance is at the minimum and hence* $avar(\hat{\theta}(S^{-1})) = (\Gamma' S^{-1} \Gamma)^{-1}$.

Proposition 3.4. ***[The Efficient GMM Estimator]*** *The GMM estimator is efficient if the weight matrix* W *is optimal, i.e.*

$$\hat{\theta}(S^{-1}) = \arg\min_{\theta} \; T \cdot f_T(\theta)' S^{-1} f_T(\theta). \tag{3.9}$$

3.3.1 Efficient GMM Estimation Procedure

To compute an efficient GMM estimator according to (3.9), the estimated matrix $\hat{S}$ of S needs to be computed. Recall that

$$\hat{S} = \mathbb{E}[\psi_t \psi_t' \hat{\varepsilon}_t^2] = \frac{1}{T} \sum_{t=1}^{T} \psi_t \psi_t' (y_t - x_t' \hat{\theta})^2. \tag{3.10}$$

Thus, we can see that a consistent estimate of θ is also required. In practice, the following two procedures are commonly used to compute the GMM estimator.

3.3.1.1 Two-Step Efficient GMM Procedure

1. Approximate $\hat{\theta}(W)$ by using an initial weight matrix $\hat{W} = I$, i.e.

$$\hat{\theta}(\hat{W}) = \arg\min_{\theta} \; T \cdot f_T(\theta)' f_T(\theta). \tag{3.11}$$

   Hence, the optimal weight matrix $\hat{S}(\hat{W})$ can be estimated from (3.10).

2. Compute the two-step efficient GMM estimator according to (3.9). That is,

$$\hat{\theta}(\hat{S}^{-1}(\hat{W})) = \arg\min_{\theta} \; T \cdot f_T(\theta)' \hat{S}^{-1}(\hat{W}) f_T(\theta). \tag{3.12}$$

Note that the asymptotic distribution of $\hat{\theta}(\hat{S}^{-1}(\hat{W}))$ is not affected by the initial weight matrix $\hat{W}$ (Imbens, 2002).

3.3.1.2 Iterated Efficient GMM Procedure

The iterated efficient GMM procedure is carried out by repeating the two-step GMM procedure until the difference of $\hat{\theta}(\hat{S}^{-1}(\hat{W}))$ from two adjacent iterations is within some pre-specified value. It is noted that from the second iteration onwards, $\hat{\theta}(\hat{S}^{-1}(\hat{W}))$ achieved from the previous iteration will be immediately used for estimating $\hat{S}(\hat{W})$ from (3.10).

3.3.2 Misspecification Test and Diagnostics

For the over-identified setting, Hansen (1982) proposed $J - statistic$ for a misspecification test of the efficient GMM estimator by defining

$$J = T \cdot f_T(\hat{\theta}(\hat{S}^{-1}))'\hat{S}^{-1}f_T(\hat{\theta}(\hat{S}^{-1})), \tag{3.13}$$

where $\hat{\theta}(\hat{S}^{-1})$ is the estimated efficient GMM estimator of θ and $\hat{S}$ is a consistent estimate of S. Under some regularity conditions and the null hypothesis that the moment function $f_t(\theta)$ has zero expectation at the true parameter value θ,

$$J \xrightarrow{d} \chi^2_{(K-L)}, \tag{3.14}$$

where K and L are the number of moment functions and unknown model parameters respectively. If $K = L$, then $J = 0$, and if $K > L$ then $J > 0$. A large J value indicates a high level of model misspecification.

If the model was rejected by the $J - statistic$, it would be possible to diagnose which moment conditions possibly cause the misspecification by investigating values of each element of the normalised moments in which we know that

$$\sqrt{T} \cdot f_T(\hat{\theta}(\hat{S}^{-1})) \xrightarrow{d} N(0, S - \Gamma(\Gamma' S^{-1}\Gamma)^{-1}\Gamma'), \text{ where } \Gamma = \mathbb{E}(\psi_t x_t).$$

For a well specified model, the individual elements (called t-ratios)

$$t_i = \frac{f_T(\hat{\theta}(\hat{S}^{-1}))_i}{([\hat{S} - \hat{\Gamma}(\hat{\Gamma}'\hat{S}^{-1}\hat{\Gamma})^{-1}\hat{\Gamma}']/T)_{ii}^{1/2}} \tag{3.15}$$

are asymptotically standard normal. Hence, the t-ratios can be used to determine the misspecification associated to each i-th moment.

3.4 Efficient Method of Moments

The efficient method of moments (EMM) is a GMM-type estimator in which the moment conditions and the weight matrix are constructed by making use of an auxiliary model. Furthermore, the methodology is also incorporated with the Monte Carlo simulation. To begin with, we first recall the properties of stationarity and ergod-

icity. Suppose that observed data y_t, for $t = -L+1, \ldots, T$, are generated from a stationary process of a model with an unknown parameter vector θ. Then, the time invariant stationary density of y_t given θ can be defined by

$$p(y_{t-L}, \ldots, y_{t-1}, y_t \mid \theta) \tag{3.16}$$

for some positive lagged value L. Moreover, if the process is also ergodic, the expectation of a time invariant function $g(y_{t-L}, ..., y_{t-1}, y_t)$ with respect to (3.16) can be approximated by

$$\mathbb{E}_\theta[g] = \frac{1}{T}\sum_{t=1}^{T} g(\hat{y}_{t-L}, \ldots, \hat{y}_{t-1}, \hat{y}_t), \tag{3.17}$$

from a large enough length T of simulation, where $\hat{y}_t$ denotes the simulated data from the model.

3.4.1 EMM Estimation Procedure

To implement EMM, we follow the two steps described by Gallant and Tauchen (2001).

3.4.1.1 The Projection Step

First, the density of observed data is supposed to be unknown. Thus, the main idea is that we approximate the transition density by projecting the data onto an auxiliary model in which its density is known. Clearly, a choice of the auxiliary models is substantially important as it needs to be able to adequately capture features of the data. Gallant and Tauchen (1989) proposed the semi-nonparametric (SNP) conditional density models to serve this purpose. The SNP models will be discussed in some detail in next section.

The observed data y_t, for $t = 1, \ldots, T$, are fitted by an SNP model, with an unknown parameter vector ρ, using the quasi maximum likelihood method. As a result, we obtain the transition density, called the score generator, $g(y_t|x_{t-1}, \tilde{\rho})$, where $x_{t-1} = (y'_{t-1}, ..., y'_{t-L})$, for some positive L and $\tilde{\rho}$ is the quasi maximum likelihood

estimator of ρ. Hence, the optimal choice of moment functions for EMM turns to be

$$\tilde{f}_t(x_{t-1}, y_t) = \frac{\partial}{\partial \rho} \log g(y_t \mid x_{t-1}, \tilde{\rho}), \tag{3.18}$$

where

$$\begin{aligned} \tilde{\rho} &= \arg\min_{\rho} \ s_T(\rho), \\ s_T(\rho) &= -\frac{1}{T} \sum_{t=1}^{T} \log g(y_t \mid x_{t-1}, \rho), \\ x_{t-1} &= (y_{t-1}, ..., y_{t-L}), \end{aligned}$$

for some positive L. Eventually, we also obtain the weight matrix $\tilde{S}$ for the GMM estimator in the form

$$\begin{aligned} \tilde{S} &= \mathbb{E}[\tilde{f}_t(x_{t-1}, y_t)\tilde{f}_t'(x_{t-1}, y_t)] \\ &= \frac{1}{T} \sum_{t=1}^{T} \tilde{f}_t(x_{t-1}, y_t)\tilde{f}_t'(x_{t-1}, y_t) \\ &= \frac{1}{T} \sum_{t=1}^{T} \left[\frac{\partial}{\partial \rho} \log g(y_t \mid x_{t-1}, \tilde{\rho}) \right] \left[\frac{\partial}{\partial \rho} \log g(y_t \mid x_{t-1}, \tilde{\rho}) \right]'. \end{aligned} \tag{3.19}$$

3.4.1.2 The Estimation Step

From the first step, we achieved the moment function $\tilde{f}_t(x_{t-1}, y_t)$ and the weight matrix $\tilde{S}$ by which it is assumed that the true unknown density is closely approximated by $g(y_t|x_{t-1}, \tilde{\rho})$. The second step is to simulate $\hat{y}_t, t = 1, \ldots, T$, directly from the original model, with an unknown parameter vector θ. Accordingly, the EMM estimator $\hat{\theta}$ of θ can be obtained from

$$\hat{\theta} = \arg\min_{\theta} \ m(\theta, \tilde{\rho})' \tilde{S}^{-1} m(\theta, \tilde{\rho}), \tag{3.20}$$

where

$$m(\theta, \tilde{\rho}) = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial}{\partial \rho} \log g(\hat{y}_t \mid \hat{x}_{t-1}, \tilde{\rho}). \tag{3.21}$$

Theorem 3.5. ***[Asymptotic Properties of the EMM Estimator]*** *Under some regularity conditions and given $\tilde{\rho}$ is the unique quasi maximum likelihood estimator of ρ, the EMM estimator $\hat{\theta}$ of θ has the following asymptotic properties:*

1. $\hat{\theta} = \theta$ *as* $T \to \infty$.
2. $\sqrt{T}\ (\hat{\theta} - \theta) \xrightarrow{d} N(0, [M\mathcal{I}^{-1}M]^{-1})$.
3. $\hat{M} = M$ *as* $T \to \infty$, *where* $\hat{M} = M(\hat{\theta}, \tilde{\rho}), M = M(\theta, \rho) = \partial(m(\theta, \rho))/\partial\theta$.
4. $\tilde{S} = \mathcal{I}$ *as* $T \to \infty$, *where*

$$\mathcal{I} = \mathbb{E}_\theta \left[\frac{\partial}{\partial \rho} \log g(y_t \mid x_{t-1}, \rho) \right] \left[\frac{\partial}{\partial \rho} \log g(y_t \mid x_{t-1}, \rho) \right]'.$$

3.4.2 Misspecification Test and Diagnostics

Similar to the GMM approach, EMM also has a $J - statistic$ for misspecification test defined by

$$J_0 = T\ m(\hat{\theta}, \tilde{\rho})' \tilde{S}^{-1} m(\hat{\theta}, \tilde{\rho}), \tag{3.22}$$

which is asymptotically chi-squared on $(p_\rho - p_\theta)$ degrees of freedom, under the null hypothesis that $p(y_{t-L}, \ldots, y_t|\theta)$ is the correct model.

Further, we also have the individual element t-ratios

$$T_i = D_i^{-1} \sqrt{T}\ m(\hat{\theta}, \tilde{\rho}), \tag{3.23}$$

where $D_i = \left(\tilde{S} - \hat{M}[\hat{M}'\tilde{S}^{-1}\hat{M}]^{-1}\hat{M}' \right)_{ii}^{1/2}$, which can be used for the diagnostic of misspecification with respect to each i-th moment.

3.5 SNP Conditional Density Models

As mentioned previously, EMM requires an auxiliary model to estimate the transition density of observed data. Generally, one may use any specific model to serve this purpose as long as it can be sufficiently fitted by the data and its density is known. We consider here the semi-nonparametric (SNP) models, introduced by Gallant and Tauchen (1989). The SNP model is known as a general purpose model that nests

several parametric and nonparametric models, from a Gaussian vector autoregressive (VAR) model to, the most general form, a nonlinear, non-Gaussian multivariate GARCH model. The SNP model parameters are estimated using maximum likelihood.

3.5.1 SNP Methodology

The central concept of SNP methodology is the use of an expansion in Hermite functions to estimate the one-step-ahead conditional density of observed data. The method lies in between parametric and nonparametric procedures. Specifically, a leading term in the SNP density function is the parametric (i.e. the Gaussian) can be extended into a nonparametric setting by an expansion in Hermite functions.

Initially, we assume that the multivariate observed data y_t, for $t = -L+1, \ldots, T$, where y_t of each t is a vector of length M and L is a number of lagged values of y_t, are stationary and ergodic. Also, we suppose the Markov property to y_t up to lagged number L and denote these lagged values by $x_{t-1} = (y'_{t-1}, ..., y'_{t-L})$, where x_{t-1} is a $M \times L$ matrix. Henceforth, the subscript of x and y may be suppressed where there is no ambiguity. Then, the SNP model is

$$y_t = \mu_{x_{t-1}} + R_{x_{t-1}} \, z_t, \tag{3.24}$$

where

(i) z_t is a random term with the density function

$$h_K(z|x) = \frac{[\mathcal{P}(z,x)]^2}{\int [\mathcal{P}(z,x)]^2 \phi(u) \, du} \, \phi(z), \tag{3.25}$$

where $\mathcal{P}(z,x)$ is a polynomial in (z,x) of degree K and $\phi(z) \sim N_M(0, I_M)$. Alternatively, $\mathcal{P}(z,x)$ can be written as

$$\mathcal{P}(z,x) = \sum_{\alpha=0}^{K_z} \left(\sum_{\beta=0}^{K_x} a_{\beta\alpha} x^\beta \right) z^\alpha, \tag{3.26}$$

where α and β are multi-indexes of maximal degree K_z and K_x respectively, and $K = K_z + K_x$. To obtain a unique representation, it is suggested to

impose a_{00} equal to one. The crucial interpretation of (3.25) is that it can be thought of a series expansion that the leading term is the normal density $\phi(z)$ and the higher order terms induce departures from normality by the function $[\mathcal{P}(z,x)]^2/\int[\mathcal{P}(z,x)]^2\phi(u)\,du$.

(ii) $\mu_{x_{t-1}}$ is the location function in a linear form of

$$\mu_{x_{t-1}} = c_0 + Cx_{t-1}, \tag{3.27}$$

where c_0 and C is a vector and a matrix of constants.

(iii) $R_{x_{t-1}}$ is the scale or volatility function and is an upper triangular matrix which depends on x_{t-1}. In fact, one can let $R_{x_{t-1}}$ be independent of x_{t-1} but it is evident that having x_{t-1} in $R_{x_{t-1}}$ can reduce the degree K_x required for achieving a sufficient approximation of the SNP transition density. $R_{x_{t-1}}$ can be defined to be consistent with the ARCH (Engle, 1982) and GARCH models as follows.

ARCH-like Specification

$$vech(R_{x_{t-1}}) = b_0 + \sum_{i=1}^{L_r} P_{(i)}|y_{t-1-L_r+i} - \mu_{x_{t-2-L_r+i}}|, \tag{3.28}$$

where $vech(R)$ is a vector of length $M(M+1)/2$ containing the elements of the upper triangle of R, b_0 is a vector of constants of length $M(M+1)/2$ and $P_{(1)}$ through $P_{(L_r)}$ are $M(M+1)/2$ by M matrices. Note that the ARCH form above is different to the classical ARCH model that defines lagged residuals in the squared form.

GARCH-like Specification

$$\begin{aligned} vech(R_{x_{t-1}}) = b_0 &+ \sum_{i=1}^{L_r} P_{(i)}|y_{t-1-L_r+i} - \mu_{x_{t-2-L_r+i}}| \\ &+ \sum_{i=1}^{L_g} diag(G_{(i)})R_{x_{t-2-L_g+i}}, \end{aligned} \tag{3.29}$$

where $G_{(1)}$ through $G_{(L_g)}$ are vectors of length $M(M+1)/2$.

With a combination of L_u, L_g, L_r, L_p and the degree K_z, K_x of the polynomial

$\mathcal{P}(z, x)$, where L_u and L_p are the lag lengths of x in μ_x and $\mathcal{P}(z, x)$ respectively, the SNP models nest considerable models as shown in Table 3.1.

Lag and Polynomial Degree Setting	Model
$L_u = 0, L_g = 0, L_r = 0, L_p \geq 0, K_z = 0, K_x = 0$	i.i.d. Gaussian
$L_u > 0, L_g = 0, L_r = 0, L_p \geq 0, K_z = 0, K_x = 0$	Gaussian VAR
$L_u > 0, L_g = 0, L_r = 0, L_p \geq 0, K_z > 0, K_x = 0$	semi-parametric VAR
$L_u \geq 0, L_g = 0, L_r > 0, L_p \geq 0, K_z = 0, K_x = 0$	Gaussian ARCH
$L_u \geq 0, L_g = 0, L_r > 0, L_p \geq 0, K_z > 0, K_x = 0$	semi-parametric ARCH
$L_u \geq 0, L_g > 0, L_r > 0, L_p \geq 0, K_z = 0, K_x = 0$	Gaussian GARCH
$L_u \geq 0, L_g > 0, L_r > 0, L_p \geq 0, K_z > 0, K_x = 0$	semi-parametric GARCH
$L_u \geq 0, L_g \geq 0, L_r \geq 0, L_p \geq 0, K_z > 0, K_x > 0$	nonlinear nonparametric

Table 3.1: The nested structure of the SNP models.

3.5.2 SNP Model Selection

It can be noticed that the SNP methodology nests a considerable number of models and hence a natural question arises as to which models should be selected. As the parameter vector ρ of the SNP models is estimated using the maximum likelihood procedure, Gallant and Tauchen (2001) suggest the model selection strategy by minimising the Bayesian Information Criterion (BIC) value such that

$$BIC(n_\rho) = s_T(\rho) + \frac{n_\rho}{2T} \log(T), \tag{3.30}$$

where $s_T(\rho)$ is the average log-likelihood in (3.18), and n_ρ is the number of parameters in the SNP model. Furthermore, for univariate models, it is recommended to search the model in the following order.

1. Determine best VAR order L_u.
2. Determine best ARCH and GARCH orders L_r, L_g.
3. Determine best z polynomial order K_z (start at $K_z = 4$).
4. Determine the best x polynomial order K_x.

For each step of the expansion, the parameter estimates of the previous fitted model should be used as starting values for fitting the next model. For multivariate models, further research on model selection still needs to be carried out.

3.6 Estimating a One-Factor CKLS Model using GMM and EMM with Daily 3-Month UK Repo Rates

GMM and EMM can be conveniently implemented to a certain number of stochastic term structure models using Finmetrics module in S-PLUS. The purpose of this section is mainly to provide a clearer picture of how each step of the GMM and EMM methodologies proceeds. The results may not be achieved nicely since it is known that a one-factor model is unlikely to capture all features of interest rate data. Here, we follow Wang and Zivot (2005), in which more examples for other applications can also be found. In this section, we estimate a one-factor CKLS model (Chan et al., 1992) with daily 3-month UK Repo rates using both GMM and EMM for comparison.

3.6.1 CKLS Model

Recall that the one-factor CKLS short rate model, given a parameter vector $\theta = (\alpha, \beta, \sigma, \gamma)'$, is defined by

$$dr_t = (\alpha + \beta r_t)dt + \sigma r_t^{\gamma} dW_t, \tag{3.31}$$

where W_t is a Wiener process. The drift term in (3.31) can be reparameterised as $k(\mu - r_t)dt$, where $\alpha = k\mu$ and $\beta = -k$, in which μ is known as the long-term mean and k the speed of mean reversion. The CKLS model is one example where we may not be able to make a strong distributional assumption on r_t since, for instance, if $\gamma = 0$, the increments in discrete time of interest rates are normally distributed, but if $\gamma = 0.5$, they have a non-central chi-squared distribution.

3.6.2 Data

Figure 3.1 shows the movement of daily 3-month UK Repo rates from 4 January 2000 to 30 June 2009. As can be seen, from mid-2008 the rates dramatically decline from approximately 5% to just above zero due to the credit crisis. Therefore, we particularly consider two periods of the data for estimating the model.

- Period A: 4 January 2000 to 30 June 2007 (1,888 observations).
- Period B: 4 January 2000 to 30 June 2009 (2,394 observations).

Period A represents a stable period, whereas Period B extends Period A to include a period of sharply declined rates.

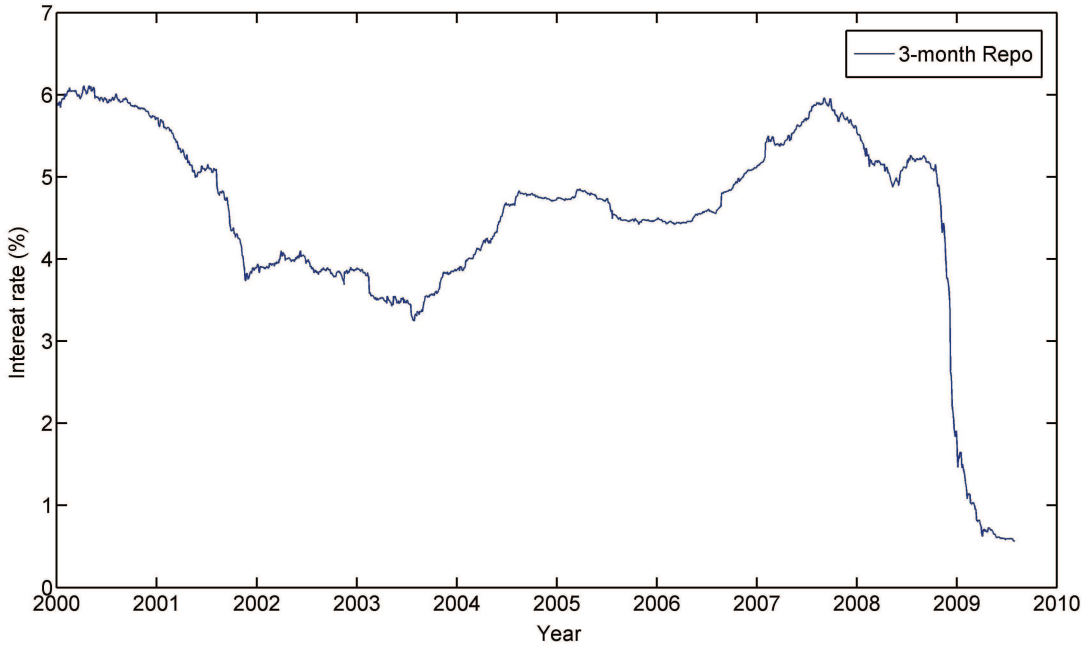

Figure 3.1: Daily 3-month UK Repo rates from 4 January 2000 to 30 June 2009.

3.6.3 Results by GMM

For the GMM estimation, we follow Chan et al. (1992), where they used the Euler approximation to discretise the process (3.31) as

$$\begin{aligned} r_{t+\Delta t} - r_t &= (\alpha + \beta r_t)\Delta t + \sigma r_t^{\gamma}\sqrt{\Delta t}\varepsilon_{t+\Delta t}, \text{ with} \\ \mathbb{E}[\varepsilon_{t+\Delta t}] &= 0,\ \mathbb{E}[\varepsilon_{t+\Delta t}^2] = 1. \end{aligned} \tag{3.32}$$

Then, the moment function $f_t(w_{t+\Delta t}, \theta)$ which is defined by

$$f_t(w_{t+\Delta t}, \theta) = \begin{pmatrix} \varepsilon_{t+\Delta t} \\ \varepsilon_{t+\Delta t} r_t \\ \varepsilon_{t+\Delta t}^2 - \sigma^2 r_t^{2\gamma}\Delta t \\ (\varepsilon_{t+\Delta t}^2 - \sigma^2 r_t^{2\gamma}\Delta t) r_t \end{pmatrix}, \tag{3.33}$$

where $w_{t+\Delta t} = (r_{t+\Delta t} - r_t, r_t, r_t^2)'$, is required to satisfy $\mathbb{E}[f_t(w_{t+\Delta t}, \theta)] = 0$.

The parameters estimated by GMM for the Repo rates over both two periods are demonstrated in Table 3.2. From the results, we obtain the long-term mean 0.0461 with speed of mean reversion 0.0847 per day for the Repo rates over Period A. Since the number of model parameters and moment conditions are equal (i.e. just-identified case), in this case the misspecification model test is not applicable. Furthermore, when considering the t values in the table, γ is found to be most significant. Chan et al. (1992) also noticed this on their results, where they used GMM to estimate the CKLS model with the annualised one-month US Treasury Bill yields from June 1964 to December 1984. Regarding the GMM results for the data over Period B, we get the higher long-term mean with negative speed of mean reversion.

GMM - Period A (1,888 observations)

	Value	Std. Error	t value	$Pr(>\|t\|)$
α	0.0039	0.0078	0.5039	0.6144
β	-0.0847	0.1628	-0.5202	0.603
σ	0.0233	0.0266	0.8785	0.3798
γ	0.6642	0.3821	1.7384	0.0823

$(\mu = -\alpha/\beta = 0.0461, k = -\beta = 0.0847)$

GMM - Period B (2,394 observations)

	Value	Std. Error	t value	$Pr(>\|t\|)$
α	-0.0250	0.0087	-2.8862	0.0039
β	0.4322	0.1680	2.5730	0.0101
σ	0.0104	0.0228	0.4577	0.6472
γ	0.2862	0.7537	0.3797	0.7042

$(\mu = -\alpha/\beta = 0.0578, k = -\beta = -0.4322)$

Table 3.2: GMM estimated parameters of one-factor CKLS model using daily 3-month UK Repo rates over Period A (top) and Period B (bottom). $Pr(>|t|)$ is the p-value with respect to the t value of each parameter.

3.6.4 Results by EMM

According to the EMM estimation, the first step is to fit the SNP models to the 3-month UK Repo rates. We first note that for the model selection, we use the SNP auto-search function in S-PLUS to search for the model that minimises the BIC values as described in the previous section. We narrow the search by constraining

the maximum number of L_p and K_x to one. In other words, the Hermite function term $\mathcal{P}(z,x)$ is not allowed to have the lagged length and the power of x_{t-1} greater than one. This is a fair compromise in order to reduce the complexity of parameter interactions in $\mathcal{P}(z,x)$.

For the Repo rates over Period A, the program suggested a semi-parametric AR(1)-GARCH(1,1) model with $K_x = 0, K_z = 8$. Specifically, from (3.24) to (3.29), we come up with the auxiliary model

$$\begin{aligned} y_t &= \mu_{x_{t-1}} + R_{x_{t-1}} z_t, \quad \text{where} \\ \mu_{x_{t-1}} &= c_0 + C x_{t-1}, \\ R_{x_{t-1}} &= (b_0 + P_1 |y_{t-1} - \mu_{x_{t-2}}| + G_1 R_{x_{t-2}}), \end{aligned} \tag{3.34}$$

and the density of z_t is $h_8(z|x) = ([\mathcal{P}(z,x)]^2 / \int [\mathcal{P}(z,x)]^2 \phi(u)du)\phi(z)$, where $\mathcal{P}(z,x) = 1 + \sum_{\alpha=1}^{8} a_\alpha z^\alpha$ and $\phi(z) \sim N(0,1)$. The parameter estimates by maximum likelihood (achieved at minimum BIC = -2.4976) are

$$\begin{aligned} c_0 = -0.0006, C = 0.9997, b_0 = 0.0071, P_1 = 0.4377, G_1 = 0.6581, \\ a_1 = 0.0590, a_2 = -0.6564, a_3 = -0.0322, a_4 = 0.1762, \\ a_5 = 0.0045, a_6 = -0.0158, a_7 = -0.0002, a_8 = 0.0005. \end{aligned} \tag{3.35}$$

Figure 3.2 shows the results of some preliminary diagnostic checks for the fitted SNP model. For the standardised residuals (top left), it can be seen that the SNP model is fitted to the data rather poorly since there still exist extreme values on several days at which the corresponding conditional standard deviations (top right) are also high. Furthermore, we can notice from the empirical distribution and the density of standardised residuals (middle left and right) that they clearly have the fatter tails on both sides. Nevertheless, almost all autocorrelation functions for the standardised residuals and particularly squared standardised residuals (bottom left and right) remain within the 95% confidence interval.

The simulated data from the fitted SNP model using the data over Period A is provided in Figure 3.3 (solid line). As can be seen, the generated data is dynamically stable. Moreover, there is no evidence of an explosion. Thus, the EMM estimation can proceed to the next step.

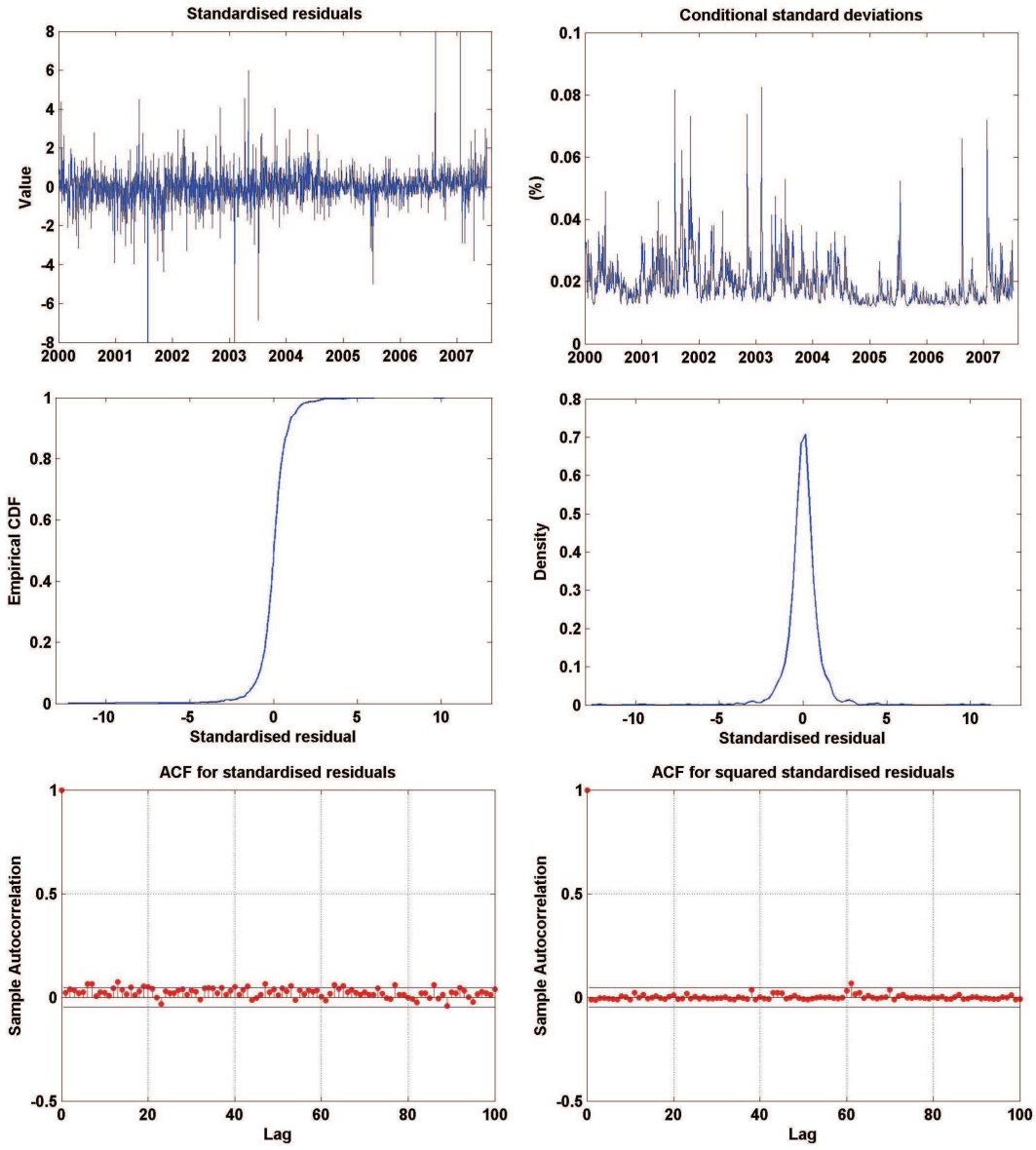

Figure 3.2: Standardised residuals (top left), conditional standard deviations (top right), empirical CDF (middle left) and density (middle right) of standardised residuals, the autocorrelation functions for the standardised (bottom left) and the squared standardised residuals (bottom right) of the fitted SNP model (AR(1)-GARCH(1,1), $K_x = 0, K_z = 8$) with daily 3-month UK Repo rates over Period A.

For the Repo rates over Period B, it turns out that the program suggested a semi-parametric AR(4)-GARCH(1,1) model with $K_x = 0, K_z = 8$ as the auxiliary model. As can be noticed from Figure 3.3 (dotted line), although the fitted SNP model can produce a sharp fall similar to the actual data, unfortunately the rates can be negative. Therefore, the next step of EMM estimation will not be carried on unless an appropriate auxiliary model can be found.

Next, we continue on to the estimation step of EMM only for the data over Period A. From the projection step, we obtained the moment conditions (or score generator)

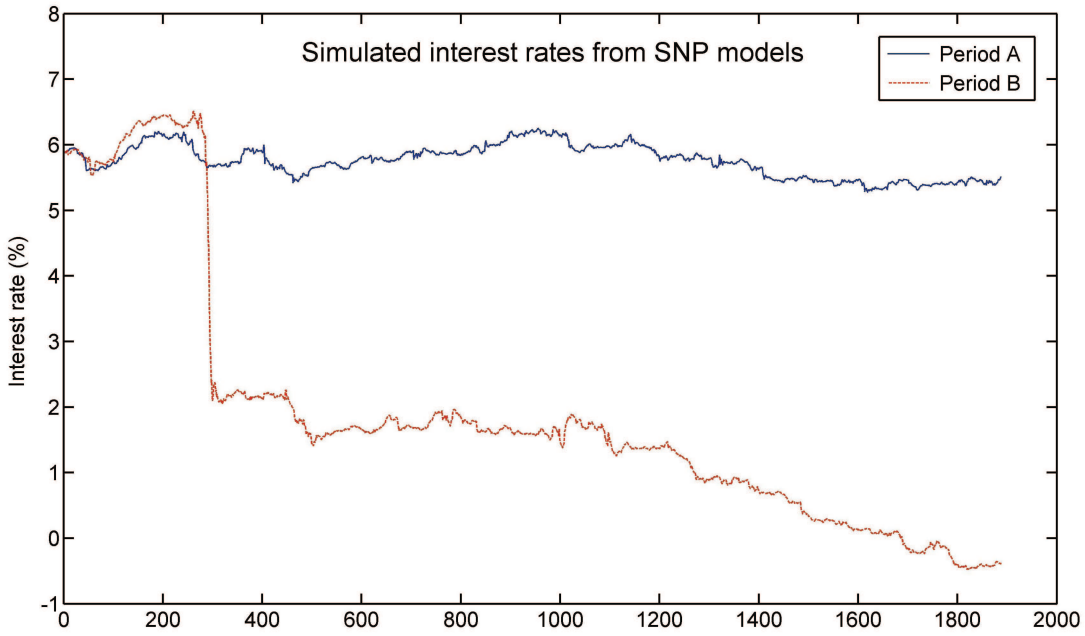

Figure 3.3: Simulated data from the fitted SNP model with daily 3-month UK Repo rates over Period A (blue, solid) and Period B (red, dotted).

EMM - Period A (1,888 observations)			
	Value	Std. Error	95% Confidence Interval
α	0.00175	0.00014	(0.00162, 0.00185)
β	-0.03725	0.00371	(-0.04104, -0.03381)
σ	2.34047	1.26179	(1.16214, 3.20102)
γ	2.43216	0.19977	(2.24033, 2.56582)
$(\mu = -\alpha/\beta = 0.04706, k = -\beta = 0.03725)$			

Table 3.3: EMM estimated parameters of one-factor CKLS model using daily 3-Month UK Repo rates over Period A. The SNP model: AR(4)-GARCH(1,1) model with $K_x = 0, K_z = 8$. EMM objective at final iteration = 40.38 (p-value = 6.488×10^{-6}).

and estimated weight matrix from the fitted SNP model. Hence, the parameters estimated by EMM can be achieved from (3.20) using simulated data from the CKLS model in (3.32). Table 3.3 presents the results of the EMM estimation with 50,000 simulated observations at a time.

The EMM objective value at final iteration (277th) is achieved at 40.38 with p-value = 6.488×10^{-6}. The degrees of freedom, which comes from the over-identification setting of the AR(1)-GARCH(1,1) model with $K_x = 0, K_z = 8$ (13 parameters) over the CKLS model (4 parameters), is equal to 9. By the EMM

approach, we have the long-term mean $\mu = 0.04706$ with speed of mean reversion $k = 0.03725$ per day.

Since the p-value is close to zero, the one-factor CKLS model is rejected using the semi-parametric AR(1)-GARCH(1,1) model with $K_x = 0, K_z = 8$ as the auxiliary model. Finally, we consider the t-ratios of each moment condition (Table 3.4). It can be noticed that the auxiliary model is likely to be misspecified with respect to the moment a_2, a_4, a_6 and a_8 as the absolute values of the t-ratios are greater than 2.

	Mean Value	Std. Error	t-ratios
a_1	-4.276	2.923	-1.463
a_2	19.359	4.874	3.972
a_3	-19.018	11.947	-1.592
a_4	128.202	34.223	3.746
a_5	-128.306	114.444	-1.121
a_6	1057.624	401.241	2.636
a_7	-1345.880	1465.161	-0.919
a_8	11713.836	5441.551	2.153
c_0	-11.042	69.888	-0.158
C	-72.153	69.219	-1.042
b_0	470.497	286.335	1.643
P_1	1.317	2.058	0.630
G_1	8.971	8.244	1.089

Table 3.4: Score information of the individual moment conditions by the semi-parametric AR(1)-GARCH(1,1) model with $K_x = 0, K_z = 8$).

As several moments related to the higher powers of z in the Hermite function were found to be misspecified, we next consider to restrict the SNP model with $K_z = 1$ for estimating the CKLS model. With the new auxiliary model, we can observe from Table 3.6 that all moment conditions are currently well specified. Moreover, according to the results presented in Table 3.5, it can be seen that the EMM estimates converge with a much smaller value of the objective function that now does not reject the overidentifying restriction with p-value = 0.0384.

Comparing to the GMM results in Table 3.2, the estimated parameters σ and γ are rather different. Recall that both GMM and EMM estimation rely on moment

functions. While the former arbitrarily defines them, the latter makes use of an auxiliary model. Accordingly, if the chosen moments from either or both methods are sub-optimal, the estimates can be distinctly different. In our results, we found that GMM estimation (in S-PLUS) for σ and γ is sensitive to initial values when the number of moments is equal to the number of parameters. Therefore, EMM is more reliable than GMM in this case.

In summary, Chapter 3 reviewed a number of modern estimation techniques in term structure modelling. Some applications of GMM and EMM with one-dimensional data were also presented in which they will be advantageous to future research with higher dimensional data. An example of estimation using MCMC with one-dimensional data can be found in Eraker (2001).

EMM - Period A (1,888 observations)			
	Value	Std. Error	95% Confidence Interval
α	0.00273	0.00136	(0.00136, 0.00409)
β	-0.06701	0.03512	(-0.10213, -0.03190)
σ	0.54409	1.49737	(-0.95328, 2.04146)
γ	1.69980	0.87642	(0.82339, 2.57622)
$(\mu = -\alpha/\beta = 0.04068, k = -\beta = 0.06701)$			

Table 3.5: EMM estimated parameters of one-factor CKLS model using daily 3-Month UK Repo rates over Period A. The SNP model: AR(4)-GARCH(1,1) model with $K_x = 0, K_z = 1$. EMM objective at final iteration = 6.52 (p-value = 0.0384).

	Mean Value	Std. Error	t-ratios
a_1	1.789	2.228	0.803
c_0	-58.210	69.125	-0.842
C	-14.577	60.002	-0.243
b_0	-459.065	624.133	-0.736
P_1	-15.939	12.574	-1.268
G_1	-18.005	16.913	-1.065

Table 3.6: Score information of the individual moment conditions by the semi-parametric AR(1)-GARCH(1,1) model with $K_x = 0, K_z = 1$).

Chapter 4

Bayesian Inference and Markov Chain Monte Carlo

Following the previous chapter, Markov Chain Monte Carlo (MCMC) methods are here discussed in more detail. We start with a brief review of Bayesian inference and then consider two main MCMC approaches: the Metropolis-Hastings (MH) algorithm and the Gibbs sampler. To implement MCMC, the popular WinBUGS software is employed to estimate the mean-reverting autoregressive AR(1) and vector autoregressive VAR(1) models with simulated data and a stochastic volatility with daily 3-month UK Repo rates. Moreover, we also re-estimate the AR(1) model using the MH methodology by coding the algorithm in Matlab for comparison with the achieved results in WinBUGS.

4.1 Bayesian Inference

It is not the intention of this section to justify the Bayesian over frequentist approach but merely to provide an introduction to its underlying concept for proceeding to the MCMC methodology in the next section. Given observed data $y = (y_1, \ldots, y_N)$, an unknown parameter vector θ of the underlying model in the Bayesian paradigm is treated as a random variable with some prior beliefs. This is contrary to the classical approach that θ is supposed to be a fixed quantity. The heart of the Bayesian approach is Bayes theorem. Here we assume that θ is continuous, and initially the joint distribution of y and θ can be written as

$$f(y,\theta) = f(y|\theta)f(\theta) = f(\theta|y)f(y).$$

Hence, it follows that the posterior distribution of θ conditional on y is

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)} = \frac{f(y|\theta)f(\theta)}{\int f(y|\theta)f(\theta)d\theta}$$

$$f(\theta|y) \propto f(y|\theta)f(\theta), \tag{4.1}$$

where $f(y|\theta)$ is the likelihood of the data, $f(\theta)$ is the prior density and $\int f(y|\theta)f(\theta)d\theta$ is the normalising constant satisfying $\int f(\theta|y)d\theta = 1$.

The posterior distribution $f(\theta|y)$ in (4.1) can also be interpreted as our prior beliefs of the parameter vector θ updated by the current information from the data. Typically, if we have little or no prior knowledge of θ, we may simply use a "non-informative" or "flat" prior which means we assign probabilities to all possible values equally likely. One popular non-informative prior is the Jeffreys prior distribution, given by

$$f(\theta) \propto [I(\theta)]^{\frac{1}{2}}, \tag{4.2}$$

where $I(\theta)$ is the Fisher information such that $I(\theta) = -\mathbb{E}_\theta[\partial^2 \log f(y|\theta)/\partial\theta\partial\theta']$.

In case that we have strong belief of the prior, it then will be called an "informative" prior. As can be noticed from (4.1), deriving the posterior in a closed-form is not trivial and most often it is hard to be achieved, especially for the parameters of a complex model. Nevertheless, there exist some priors that belong to the same class as the posterior distributions, the so-called "conjugate" prior distributions. Some conjugate distributions are presented in Table 4.1.

Prior Distribution	Likelihood of $y_1, \ldots, y_N$	Posterior Distribution
$\theta \sim \text{Normal}(a, b^2)$	Normal (θ, σ^2)	Normal $(\frac{a\sigma^2+Nb^2\bar{y}}{\sigma^2+Nb^2}, \frac{b^2\sigma^2}{\sigma^2+Nb^2})$
$\theta \sim$ Gamma (a, b)	Poisson(θ)	Gamma$(a + \sum y_i, b + N)$
$\theta \sim$ Gamma (a, b)	Exponential (θ)	Gamma$(a + N, b + \sum y_i)$
$\theta \sim \text{Beta}(a, b)$	Bernoulli(θ)	Beta$(a + \sum y_i, b + N - \sum y_i)$

Table 4.1: Some conjugate distributions.

4.2 Markov Chain Monte Carlo Methods

As mentioned before, the closed-form posterior distribution for a complex model frequently cannot be obtained analytically and hence this was a crucial shortfall in terms of practical implementation of the Bayesian estimation in the past. Fortunately, the development of MCMC methods in recent decades has enabled us to deal with such problems by achieving the posterior distribution by simulation. In this section, we describe the two main MCMC algorithms and some technical Markov chain properties are provided in Appendix B.

4.2.1 The Metropolis-Hastings Algorithm

With reference to Garmerman et al. (2006), suppose that we have observed data y and the model parameter vector $\theta = (\theta_1, \ldots, \theta_d)$ such that we cannot draw easily from the posterior distribution $\pi(\theta|y)$ for θ. The Metropolis-Hastings (MH) algorithm allows us to avoid the direct simulation from $\pi(\theta|y)$ by making use of a proposal distribution and computing the acceptance probability for a candidate point through the following procedure:

1. Initialise the chain to $\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_d^{(0)})$ and set $t = 1$.
2. Update each element of $\theta^{(t)}$ with substeps $j = 1, \ldots, d$
   - Denote $\tilde{\theta}^{(t)}$ as the latest $\theta^{(t)}$ that includes updated values of the elements from earlier substeps.
   - Now for each substep $j = 1, \ldots, d$.
   - Sample a candidate point θ^* from a proposal distribution $q(\cdot|\tilde{\theta}^{(t)})$.
   - Generate U from $\sim U[0,1]$ and compute the acceptance probability

     $$\alpha(\tilde{\theta}_{j-1}^{(t)}, \theta^*) = \min\left\{1, \frac{\pi(\theta^*|y)q(\tilde{\theta}_{j-1}^{(t)}|\theta^*)}{\pi(\tilde{\theta}_{j-1}^{(t)}|y)q(\theta^*|\tilde{\theta}_{j-1}^{(t)})}\right\}, \tag{4.3}$$

     where we define $\tilde{\theta}_0^{(t)} = \tilde{\theta}_d^{(t-1)}$.
   - If $U \leq \alpha(\tilde{\theta}_{j-1}^{(t)}, \theta^*)$, then set $\tilde{\theta}_j^{(t)} = \theta^*$, otherwise $\tilde{\theta}_j^{(t)} = \tilde{\theta}_{j-1}^{(t)}$.
3. Set $t = t + 1$ and repeat step 2 until the last iteration (convergence).

Note that the elements of θ may be grouped and updated at once, rather than a single element. It is clear that the proposal distribution $q(\cdot|\theta^{(t)})$ plays a main role in facilitating the MH algorithm. The selected proposal distribution should be easy to sample and theoretically it is also required to satisfy the irreducible and aperiodic properties of Markov chains. Irreducibility means there exists a positive probability that the chain can reach any non-empty set from all different sets of states. Aperiodicity ensures that the chain will not stick in a particular part of the state space.

Furthermore, it can be observed from (4.3) that the acceptance probability appears in terms of a ratio of the posterior distribution $\pi(\cdot|y)$ so that the normalising constant term in (4.1) can be cancelled out. This is an important advantage of the MH algorithm. Additionally, if the chosen proposal distribution is symmetric, i.e. $q(\tilde{\theta}_{j-1}^{(t)}|\theta^*) = q(\theta^*|\theta_{j-1}^{(t)})$, the acceptance probability in (4.3) will be simplified to

$$\alpha(\tilde{\theta}_{j-1}^{(t)}, \theta^*) = \min\left\{1, \frac{\pi(\theta^*|y)}{\pi(\tilde{\theta}_{j-1}^{(t)}|y)}\right\}. \tag{4.4}$$

When θ^* is assumed to be independent of $\tilde{\theta}_{j-1}^{(t)}$, the acceptance probability turns to be

$$\alpha(\tilde{\theta}_{j-1}^{(t)}, \theta^*) = \min\left\{1, \frac{\pi(\theta^*|y)q(\tilde{\theta}_{j-1}^{(t)})}{\pi(\tilde{\theta}_{j-1}^{(t)}|y)q(\theta^*)}\right\}. \tag{4.5}$$

4.2.2 The Gibbs Sampler

The Gibbs sampler can be categorised as a special case of the MH algorithm where the acceptance probability is always equal to 1 but the full conditional posterior distribution must be known and simulation from it must be easy. Therefore, a proposal distribution is not required since a sample can be drawn from the full conditional posterior distribution.

As before, consider observed data y and the posterior distribution $\pi(\theta|y)$, where $\theta = (\theta_1, \ldots, \theta_d)$ is the parameter vector of the model of interest. The Gibbs sampler methodology is as follows.

1. Set $t = 0$ and initialise the chain to $\theta^{(0)} = (\theta_1^{(0)}, \ldots, \theta_d^{(0)})$.

2. Now for $t = 1, 2, \ldots$, until convergence draw a sample for each element of θ by

$$\begin{aligned}
\theta_1^{(t)} &\sim \pi(\theta_1|y, \theta_2^{(t-1)}, \ldots, \theta_d^{(t-1)}) \\
\theta_2^{(t)} &\sim \pi(\theta_2|y, \theta_1^{(t)}, \theta_3^{(t-1)}, \ldots, \theta_d^{(t-1)}) \\
&\vdots \\
\theta_k^{(t)} &\sim \pi(\theta_k|y, \theta_1^{(t)}, \ldots, \theta_{k-1}^{(t)}, \theta_{k+1}^{(t-1)}, \ldots, \theta_d^{(t-1)}) \\
&\vdots \\
\theta_d^{(t)} &\sim \pi(\theta_d|y, \theta_1^{(t)}, \ldots, \theta_{d-1}^{(t)}).
\end{aligned}$$

4.2.3 Inference and Convergence Considerations

The ultimate goal of MCMC simulation is to make inferences from the simulated sample paths whose distributions attain stationarity and converge to, ideally, the true posterior distribution of the parameters of interest. Several questions arise as to how to judge when convergence is achieved and how long the simulation should be sufficient to be able to draw a conclusion for the convergence. To date, these are still practical issues that need further research. Nevertheless, some standard statistical concepts may be employed to preliminarily assess the convergence as described below.

1. *Sample plots.* Presumably having run the simulation long enough, one may first plot entire sample paths for each parameter and then decide a cut-off point at discretion where all the chains start stabilising and moving around some specific values. The samples before the cut-off, namely the burn-in iterations, will be discarded and the remaining samples (assuming that the distribution gets stationary) will be used for making inferences such as means, medians, variances, credible intervals and so on.

2. *Autocorrelations.* Typically, we also wish the successive draws of each chain be independent. To quantify this, the autocorrelation function (ACF) is commonly used. If high autocorrelations occur, they may be mitigated by using a *"thinning"* procedure. That is, rather than taking account for every single successive draw, we pick up the sample values only from every k-th iteration. As a consequence, the simulation needs to be run much longer.

3. *Cross-correlations.* For multi-parameter models, it is suggested that we should also check correlations across parameters to investigate the correlation struc-

ture.

To make inferences, we may also consider posterior densities of the parameters which are the probability densities computed from the sample paths. A plot of this allows us to visualise shape of distribution, data dispersion, mode, etc.

Apart from the above tools, various convergence diagnostics have been proposed by researchers. A good comparative review of some of those can be found in Cowles and Carlin (1996), where the authors tested 13 diagnostic methods on a trivariate normal (with high correlations) model. In their summary, unfortunately, each has its own pitfalls and they are fairly difficult to use. Furthermore, many of them can sometimes fail to detect convergence failure, even in low-dimensional problems. One popular convergence diagnostic which is frequently mentioned is the Gelman-Rubin approach. The method requires us to simulate several chains with different starting points simultaneously. The convergence is then measured by a scalar quantity which is computed from the variances of within a chain and between the chains.

4.3 MCMC in WinBUGS

The difficulty of implementing MCMC algorithms is much based on complexity of the model being considered. At the beginning, one may first think of using the WinBUGS software which is a Windows platform of the program BUGS (Bayesian inference Using Gibbs Samplings). Nowadays, WinBUGS is a popular software widely used by practitioners for MCMC simulation due to its convenience to use. More details can be referred to the WinBUGS user manual (Spiegelhalter et al., 2003).

In this section, we use MCMC in WinBUGS to estimate three different kinds of models: autoregressive AR(1), bivariate autoregressive VAR(1) and stochastic volatility models.

4.3.1 Autoregressive AR(1) Model

We begin with estimating a simple model by assuming that observed data $X(1)$, $X(2), \ldots, X(N)$ follow a mean-reverting AR(1) process, i.e.

$$X(i) \;=\; \mu + k(X(i-1) - \mu) + \sigma Z(i),$$

$$X(i) \sim N(\mu + k(X(i-1) - \mu), \sigma^2), \tag{4.6}$$

where $Z(i) \sim i.i.d.\ N(0,1)$. The model parameter vector is $\theta = (\mu, k, \tau)$, where τ is the precision parameter equal to $1/\sigma^2$. Implementing this model in WinBUGS is trivial. What we need to know is the syntax for defining data, models, prior distributions and initial values of parameters. The graphical model (as shown in Figure 4.1) may be drawn to provide a clearer dependency structure of the model parameters.

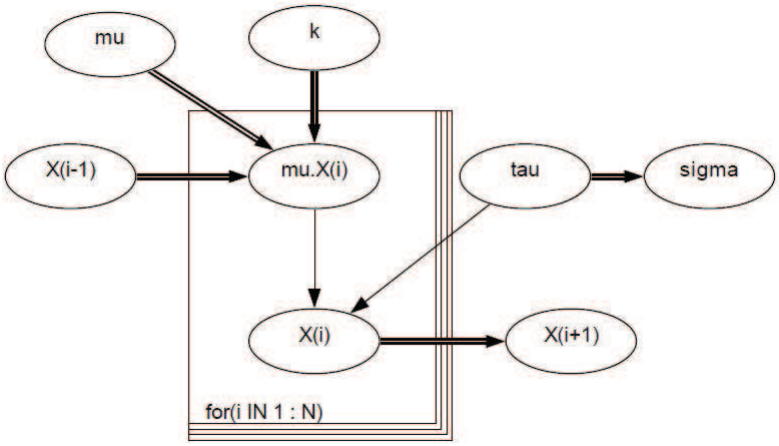

Figure 4.1: Graphical model for the mean-reverting AR(1) model.

In Figure 4.1, each ellipse indicates either a stochastic or a deterministic node which is independent of all the other nodes given its parent nodes (the nodes that arrows are pointing to). Stochastic nodes are the nodes for which the distributions are provided, whereas deterministic nodes are logical functions of other nodes. Typically, there are two kinds of arrows used to express the relationships. A solid arrow specifies the conditional independence between nodes and a hollow arrow indicates a logical function. For example, $\mathbb{E}[X(i)|X(i-1), \mu, k]$ (the mean of the normal distribution in (4.6)) is a deterministic node since it is a function comprising of the nodes μ, k and $X(i-1)$.

Next, the prior distributions are assigned as follows.

$$\begin{aligned} f(\mu): &\quad \mu \sim N(0, 1 \times 10^6), \\ f(k): &\quad k = 2\phi - 1, \text{ where } \phi \sim \text{Beta}(1,1), \end{aligned}$$

$$f(\tau): \quad \tau \sim \text{Gamma}(0.001, 100), \tag{4.7}$$

In general, non-informative priors are specified. However, in case of parameter k, we know that for a model of this kind, $|k|$ should be less than 1 in order for the process be stationary. Instead of assigning $U[-1, 1]$ to k, we use some transformation by using the beta distribution (which is more informative than the uniform distribution). It is known that $U[0, 1]$ is equivalent to $B(1, 1)$ which has values in the range of $(0, 1)$ with mean and standard deviation of 0.5 and 0.289 respectively. Additionally, the gamma prior distribution is also specified for a non-negative parameter, i.e. τ. The gamma probability density function used here is parameterised in terms of a shape parameter a and scale parameter b such that

$$f(x; a, b) = \frac{1}{b^a \Gamma(a)} x^{a-1} e^{-\frac{x}{b}},$$

where $\Gamma(\cdot)$ is the Gamma function.

4.3.1.1 Simulated Data

The data simulated from the AR(1) model in (4.6), for $N = 1500$, are plotted in Figure 4.2. The parameter values are given as $\mu = -2.0, k = 0.5, \sigma = 0.4$ with initial value $X(1) = -2.0$.

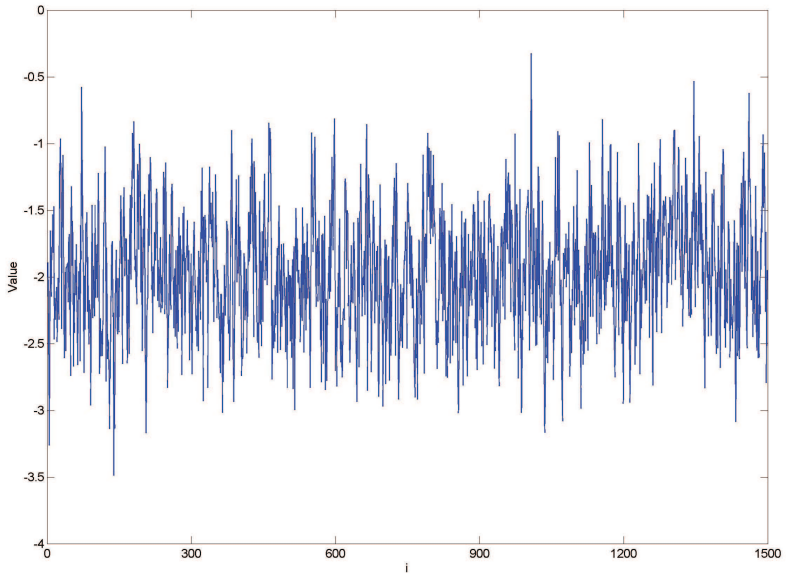

Figure 4.2: Simulated data from the mean-reverting AR(1) model, given $\mu = -2.0, k = 0.5, \sigma = 0.4$ and $X(1) = -2.0$, for $i = 1, \ldots, 1500$.

4.3.1.2 Results

We run MCMC in WinBUGS for 10,000 iterations to estimate the mean-reverting AR(1) model in (4.6) with simulated data. Regarding the first observation $X(1)$, its unconditional likelihood may be derived (will be shown in the next section) or ignored if the dataset is large. In this case, since WinBUGS does not have much flexibility to insert a complicated mathematical formula, we simply assume $X(1) \sim (\mu + k(X(0) - \mu), \sigma^2)$, where $X(0) \sim N(\mu, \sigma^2)$ is an additional parameter.

Figure 4.3 illustrates the sample paths of model parameters μ, k and σ from the 501st to the 10,000th iteration. The first 500 iterations are discarded as a burn-in period. According to the results, we can see that the chains converge very well for all parameters with the estimated means (Table 4.2) very close to the true values (within one standard deviation). Clearly, all the posterior densities (Figure 4.4, top) are smooth and have a symmetrical shape. Moreover, it can be observed from Figure 4.4 (bottom) that the autocorrelation functions all thin out rapidly even from the first lag.

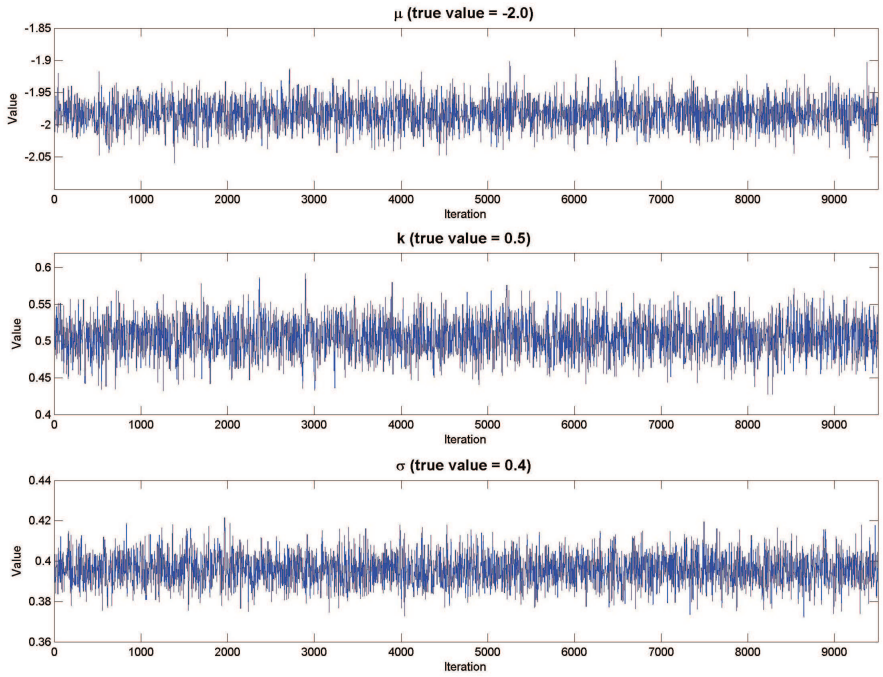

Figure 4.3: Sample paths of model parameters of the mean reverting AR(1) model using MCMC in WinBUGS.

	(True value)	Mean	Std.	95% Credible Interval	Sample
μ	(-2.0)	-1.983	0.0206	(-2.023,-1.943)	9,500
k	(0.5)	0.505	0.0223	(0.4614,0.5487)	9,500
σ	(0.4)	0.396	0.0073	(0.3817,0.4103)	9,500

Table 4.2: Summary statistics of parameter posterior estimates of the mean reverting AR(1) model using MCMC in WinBUGS (from 501st to 10,000th iteration).

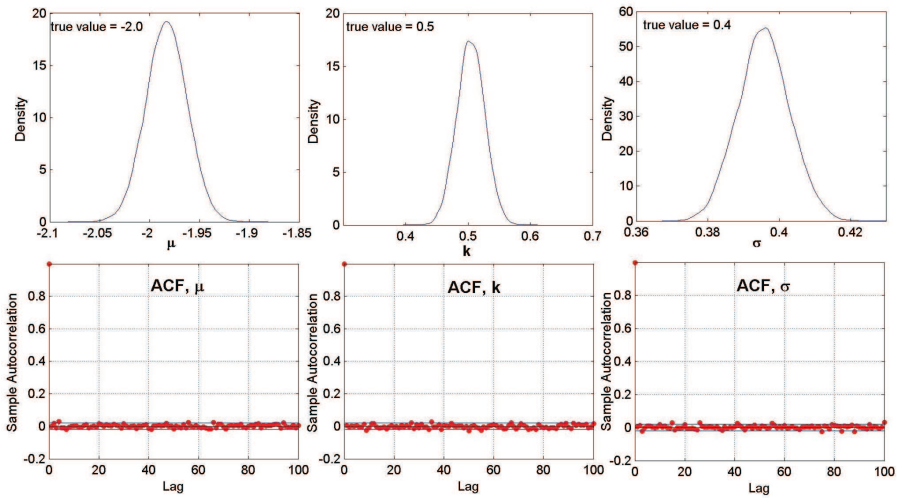

Figure 4.4: Posterior densities (top) and autocorrelation functions of model parameters of the mean reverting AR(1) model using MCMC in WinBUGS.

4.3.2 Bivariate Vector Autoregressive VAR(1) Model

We next consider estimation of the mean-reverting bivariate normal VAR(1) model in WinBUGS. Suppose that observed data $X(i) = (X_1(i), X_2(i))'$, for $i = 1, \ldots, N$, follow a VAR(1) process such that

$$(4.8)\quad \begin{pmatrix} X_1(i) \\ X_2(i) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix} \begin{pmatrix} X_1(i-1) - \mu_1 \\ X_2(i-1) - \mu_2 \end{pmatrix} + \begin{pmatrix} Z_1(i) \\ Z_2(i) \end{pmatrix}, \text{ or}$$

$$X(i) = \mu + K(X(i-1) - \mu) + Z(i),$$

where $X(0) \sim N_2(\mu, \Sigma)$, $Z(t) \sim N_2(0, \Sigma)$, with

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

being a covariance matrix.

For the prior distributions, we define

$$\begin{aligned} f(\mu): &\quad \mu_1, \mu_2 \sim N(0, 1 \times 10^6), \\ f(k): &\quad k_1, k_2 = 2\phi - 1, \text{ where } \phi \sim \text{Beta}(1,1), \\ f(\Sigma^{-1}): &\quad \Sigma^{-1} \sim \text{Wishart}(R, 2). \end{aligned} \qquad (4.9)$$

The priors for μ_1, μ_2, k_1 and k_2 are the same as previously described for the AR(1) model. For the covariance matrix Σ, its inverse Σ^{-1} is called the precision matrix and is assumed to follow a Wishart $(R, 2)$ distribution with density

$$f(x; R, k) = |R|^{k/2}|x|^{(k-p-1)/2} \exp(-\frac{1}{2}Tr(Rx)),$$

where $k = rank(R)$ and x being a positive definite $p \times p$ symmetric matrix.

Note that the Wishart distribution is a generalisation of the univariate chi-square distribution for two or more variables. It is a distribution commonly used for the covariance matrix of this kind of model.

4.3.2.1 Simulated Data

Figure 4.5 shows the simulated data $X_1(i)$ and $X_2(i)$, for $i = 1, \ldots, 1500$, from the bivariate VAR(1) model in (4.8). The parameter values are given as $\mu_1 = \mu_2 = 0, k_1 = 0.6, k_2 = 0.06, \sigma_1 = 0.6, \sigma_2 = 0.4$ and $\rho = -0.5$ with initial values $X(1) = (2, 3)'$.

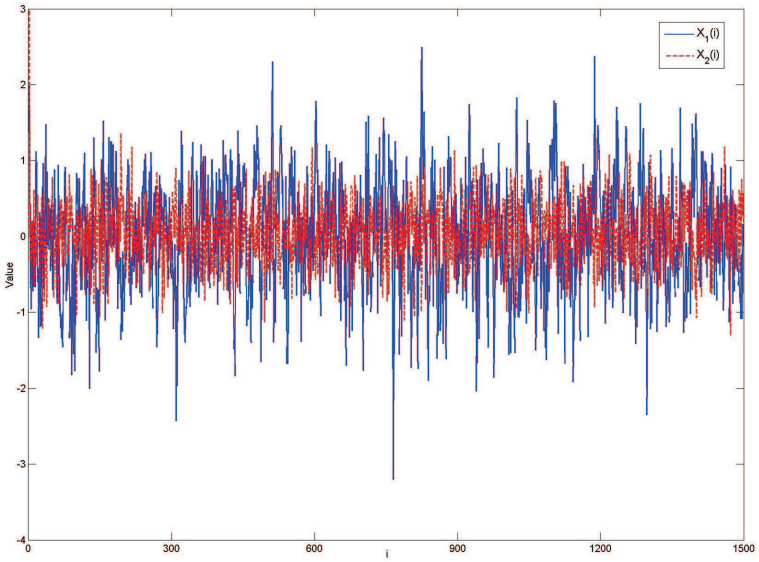

Figure 4.5: Simulated $X_1(i)$ (solid) and $X_2(i)$ (dotted) from the mean-reverting bivariate VAR(1) model, given $\mu_1 = \mu_2 = 0, k_1 = 0.6, k_2 = 0.06, \sigma_1 = 0.6, \sigma_2 = 0.4$ and $\rho = -0.5$ with initial values $X(1) = (2, 3)'$, for $i = 1, \ldots, 1500$.

4.3.2.2 Results

As before, we run 10,000 iterations of MCMC simulation in WinBUGS for estimating the mean-reverting VAR(1) model in (4.8) with simulated data. The first 500 iterations are considered as a burn-in period.

The sample paths of the simulation with summary statistics of parameter posterior estimates are provided in Figure 4.6 and Table 4.3 respectively. From the figure, we can notice that all the chains converge very well and reasonably encompass the true values, although μ_2 is found to be most biased (but its mean value is still within the 95% credible interval). According to the posterior densities shown in Figure 4.7, we can see that parameters are estimated well. Finally, it can be observed from Figure 4.8 that there is no evidence of high autocorrelation functions for all parameters after the first couple lags.

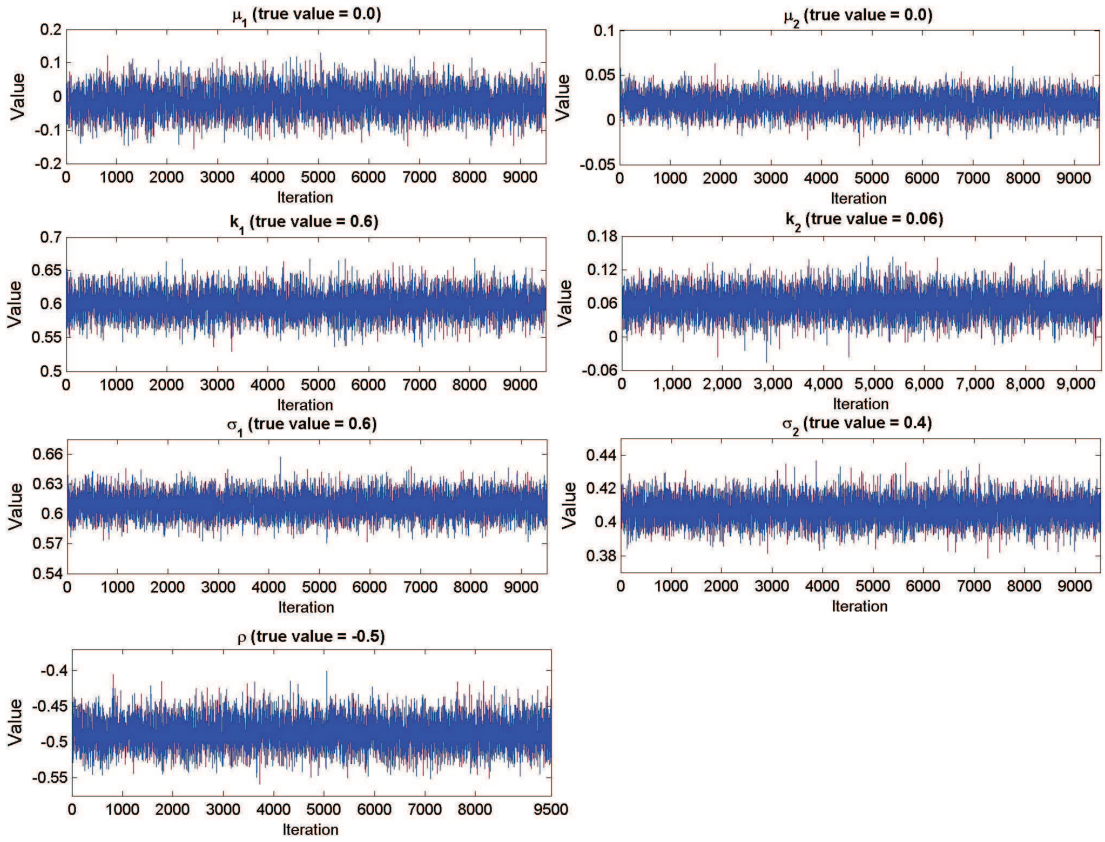

Figure 4.6: Sample paths of model parameters of the mean-reverting bivariate VAR(1) model using MCMC in WinBUGS.

	(True value)	Mean	Std.	95% Credible Interval	Sample
μ_1	(0.0)	-0.015	0.0394	(-0.0921,0.0634)	9,500
μ_2	(0.0)	0.018	0.0112	(-0.0037,0.0400)	9,500
k_1	(0.6)	0.600	0.0224	(0.5640,0.6378)	9,500
k_2	(0.06)	0.0596	0.02350	(0.01342,0.10570)	9,500
σ_1	(0.6)	0.609	0.0141	(0.5874,0.6315)	9,500
σ_2	(0.4)	0.407	0.0077	(0.3928,0.4217)	9,500
ρ	(-0.5)	-0.487	0.0203	(-0.5265,-0.4467)	9,500

Table 4.3: Summary statistics of parameter posterior estimates of the mean-reverting bivariate VAR(1) model using MCMC in WinBUGS.

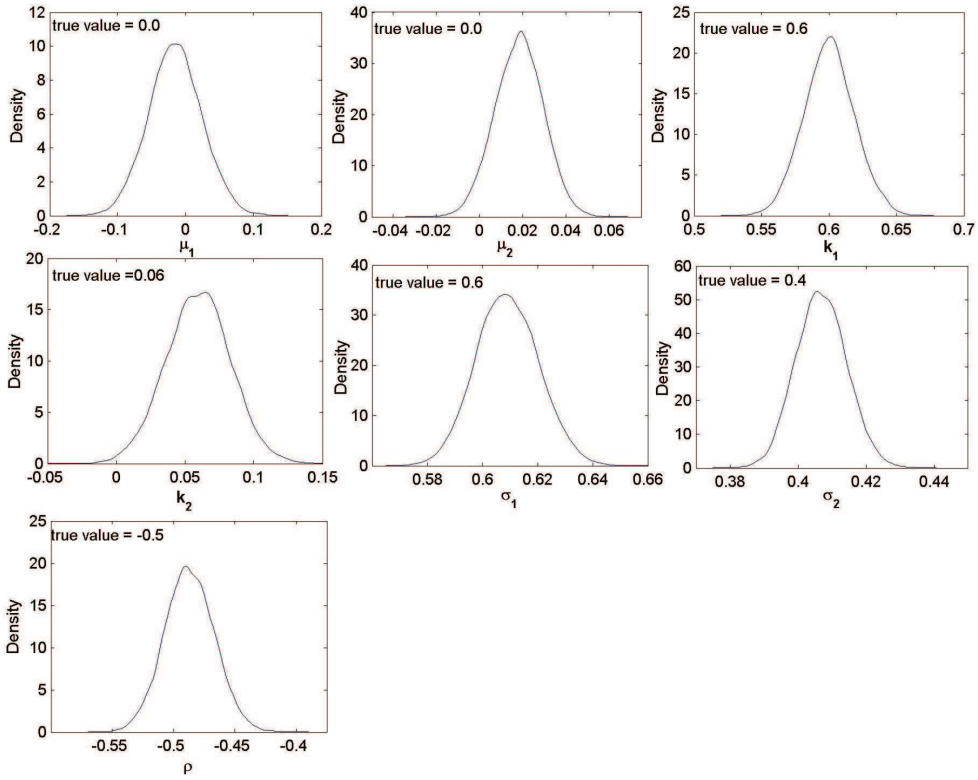

Figure 4.7: Posterior densities of model parameters of the mean-reverting bivariate VAR(1) model using MCMC in WinBUGS.

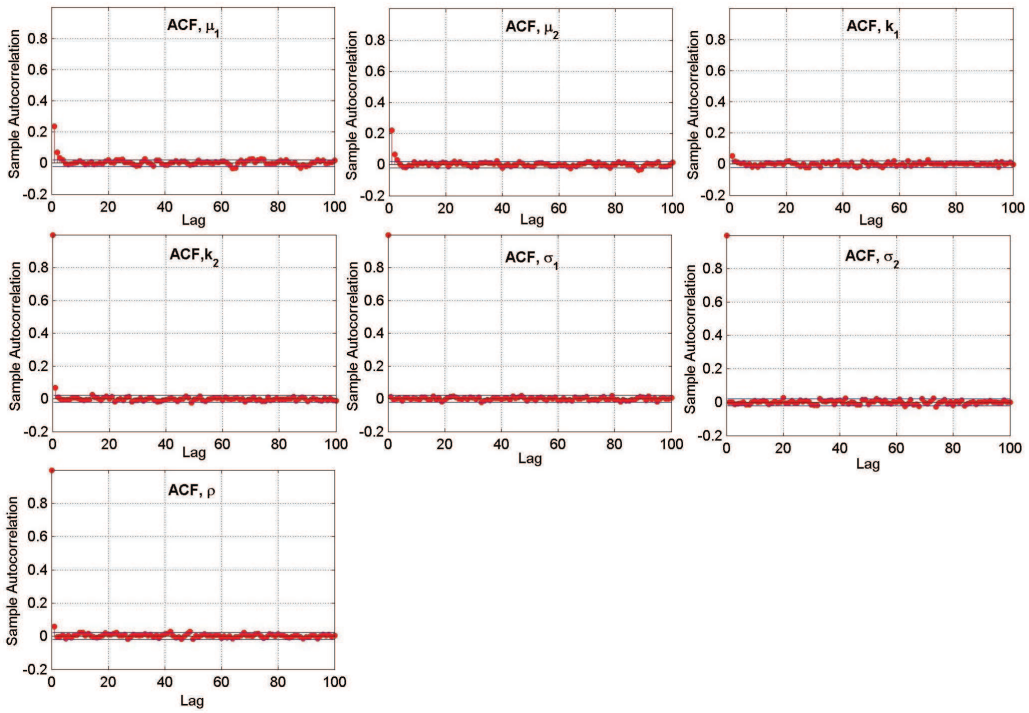

Figure 4.8: Autocorrelation functions of model parameters of the mean-reverting bivariate VAR(1) model using MCMC in WinBUGS.

4.3.3 Stochastic Volatility Models

The last model considered for illustrating the MCMC estimation in WinBUGS is a stochastic volatility (SV) model in which we follow Meyer and Yu (2000) such that, given observed data $X(i)$, for $i = 1, \ldots, N$,

$$\begin{aligned} X(i)|\theta(i) &= \exp\left(\frac{1}{2}\theta(i)\right) Z_1(i), \\ \theta(i) &= \mu + k(\theta(i-1) - \mu) + \sigma Z_2(i), \end{aligned} \tag{4.10}$$

where $Z_1(i), Z_2(i) \sim i.i.d.\ N(0,1)$, for $i = 1, \ldots, N$, $Z_1(i)$ and $Z_2(i)$ are independent, and $\theta(i)$ is the latent stochastic volatility of $X(i)$, where we assume $\theta(0) \sim N(\mu, \sigma^2)$.

From (4.10), the joint posterior distribution of all unknown quantities in this model then can be defined by

$$\begin{aligned} f(\mu, k, \sigma^2, \theta(0), \ldots, \theta(N)|X) &\propto \prod_{i=1}^{N} f(X(i)|\theta(i)) \times \\ &\prod_{i=1}^{N} f(\theta(i)|\theta(i-1), \mu, k, \sigma^2) f(\theta(0)|\mu, \sigma^2) f(\mu) f(k) f(\sigma^2), \end{aligned} \tag{4.11}$$

where the prior distributions of the parameters μ, k and σ^2 are assumed to be independent such that

$$\begin{aligned} f(\mu): \quad \mu &\sim N(0, 1 \times 10^6), \\ f(k): \quad k &= 2\phi - 1, \text{ where } \phi \sim \text{Beta}(1,1), \\ f(\tau): \quad \tau &\sim \text{Gamma}(0.001, 100), \end{aligned} \tag{4.12}$$

where τ is the precision parameter equal to $1/\sigma^2$.

4.3.3.1 Data

In contrast to our first two WinBUGS examples, we use here a real dataset. The daily yield changes of 3-month Repo UK rates from 4 January 2000 to 30 June 2007

are considered. Specifically, we define

$$X(i) = y(i) - y(i-1), \tag{4.13}$$

where $y(i)$ is an observed yield (in percentage) on day i. Daily yield changes (1,887 observations) are plotted in Figure 4.9.

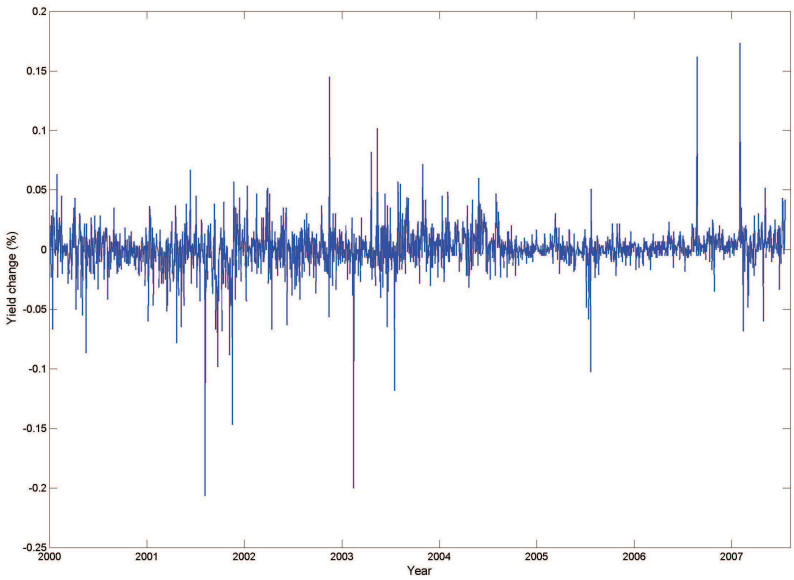

Figure 4.9: Daily yield changes of 3-month UK Repo rates from 4 January 2000 to 30 June 2007 (1,887 observations).

4.3.3.2 Results

Figure 4.10 shows sample paths of the parameters μ, k and σ (with initial values of 0.0, 0.0 and 1.0 respectively) for 20,000 iterations of MCMC simulation in WinBUGS from estimating the SV model in (4.10) with the daily yield changes of 3-month UK Repo rates (Figure 4.9). The first 2,000 iterations (left column) are chosen as a burn-in period and it can be seen that the chains do not converge to some specific values within the first 1,000 iterations. Afterwards (right column), it is obvious that all chains look stationary. The summary statistics of the parameter posterior estimates are given in Table 4.4. In Figure 4.11, autocorrelation functions of all three parameters are demonstrated, in which we can observe that those of k and σ are relatively much higher than those of μ. Nevertheless, this may be improved by using a thinning procedure where in this case we may consider to take the values of every 300-th iteration from the simulation.

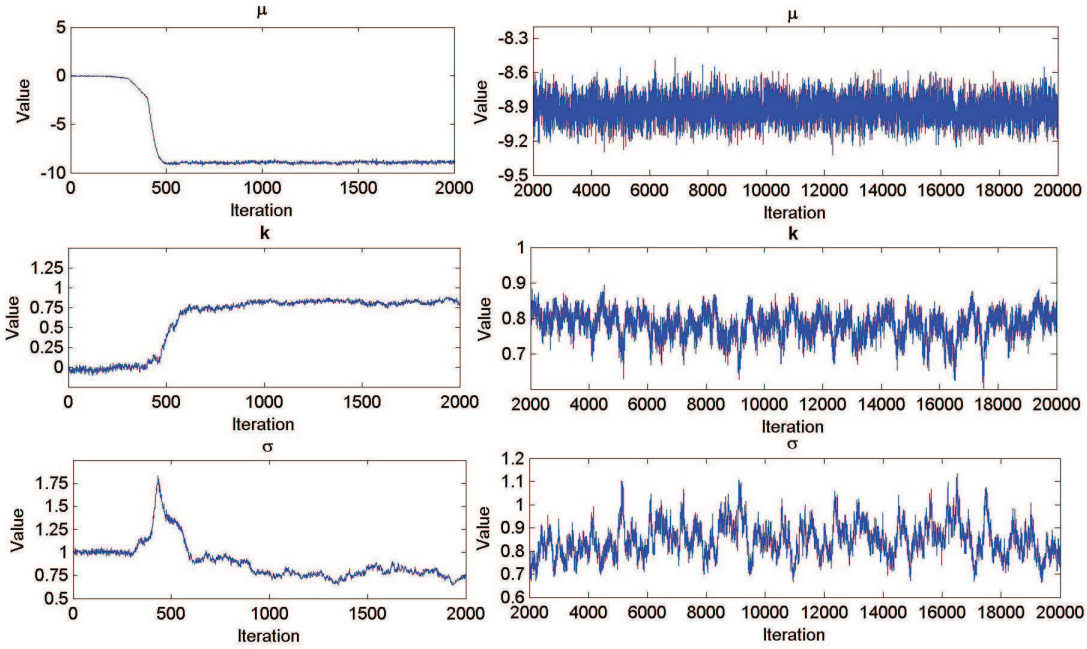

Figure 4.10: Sample paths of model parameters of the stochastic volatility model using MCMC in WinBUGS, $Z_1(i)$ and $Z_2(i)$ are independent.

	Mean	Std.	95% Credible Interval	Sample
μ	-8.914	0.1020	(-9.1120,-8.7110)	18,000
k	0.782	0.0377	(0.6962,0.8455)	18,000
σ	0.848	0.0715	(0.7208,1.0030)	18,000

Table 4.4: Summary statistics of parameter posterior estimates of the stochastic volatility model using MCMC in WinBUGS, $Z_1(i)$ and $Z_2(i)$ are independent.

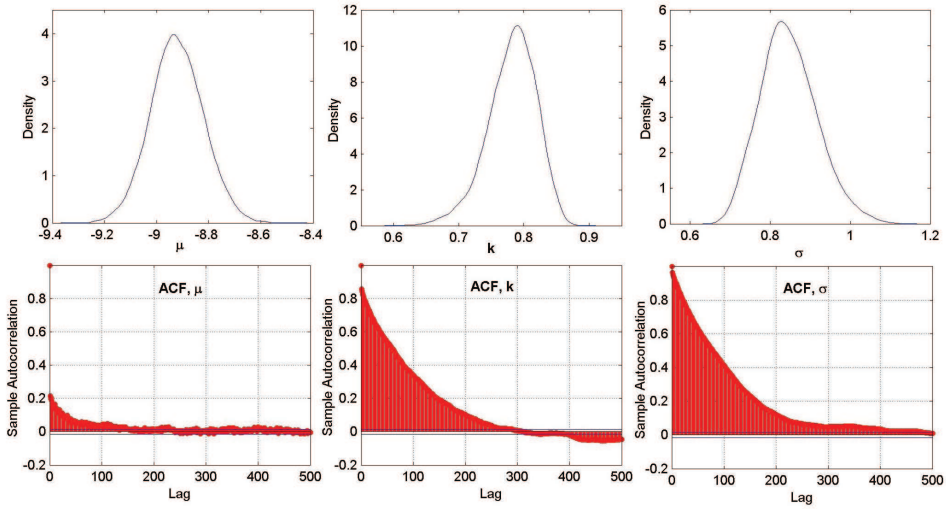

Figure 4.11: Posterior densities (top) and autocorrelation functions (bottom) of model parameters of the stochastic volatility model using MCMC in WinBUGS, $Z_1(i)$ and $Z_2(i)$ are independent.

Leverage Effect

Additionally, we investigate one more case of the SV model in (4.10), where $Z_1(i)$ and $Z_2(i+1)$ are allowed to be correlated. More precisely, with reference to Meyer and Yu (2000), we define

$$X(i)|\theta(i) = \exp\left(\frac{1}{2}\theta(i)\right) Z_1(i), \quad i = 1, \ldots, N,$$

$$\theta(i+1) = \mu + k(\theta(i) - \mu) + \sigma Z_2(i), \quad i = 1, \ldots, N-1, \tag{4.14}$$

where

$$\begin{pmatrix} Z_1(i) \\ Z_2(i+1) \end{pmatrix} \overset{i.i.d.}{\sim} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

This effect is called the "leverage effect". From (4.14), it follows that

$$\theta(i+1) \sim N(\mu + k(\theta(i) - \mu), \sigma^2),$$

$$X(i) \sim N\left[\frac{\rho}{\sigma}\exp(\theta(i)/2)(\theta(i+1) - \mu - k(\theta(i) - \mu)), \exp(\theta(i))(1-\rho^2)\right].$$

The priors of μ, k and σ are specified as in (4.12), whereas the prior of the additional parameter ρ is defined by $\rho = 2\psi - 1$, where $\psi \sim \text{Beta}(1,1)$. We estimate the SV model with leverage effect in WinBUGS with 20,000 iterations using the same data as in Section 4.3.3.1 (daily yield changes of 3-month UK Repo rates). Comparing the results of the SV model with leverage effect (Figures 4.12 and 4.13 and Table 4.5) to the previous estimates without leverage effect (Figures 4.10 and 4.11 and Table 4.4), some observations can be made as follows.

- The convergence of μ turns to be very slow, while k and σ still converge reasonably well. Parameter ρ also converges well.
- Comparing Table 4.5 to 4.4, we can observe that all parameter estimates are distinctly different. Specifically, with leverage effect, k appears to be very close to 1 (which means the stochastic volatility process $\theta(i)$ is very persistent) and σ is much smaller. The estimate of ρ is approximately equal to 0.
- The autocorrelation functions of μ with leverage effect are extremely high, whereas those of σ and particularly of k are improved. The autocorrelation

functions of ρ sharply decline within the first 50 lags.

In conclusion, the leverage effect seems unlikely to improve the overall estimation of the SV model using daily yield changes of 3-month UK Repo rates. Although the autocorrelation functions of some parameters are obviously improved, the estimation of μ seems worse. Despite obtaining the estimated mean of ρ close to 0, the other parameter estimates for both cases are distinctly different. Perhaps, this is because the poor estimate of μ with leverage effect.

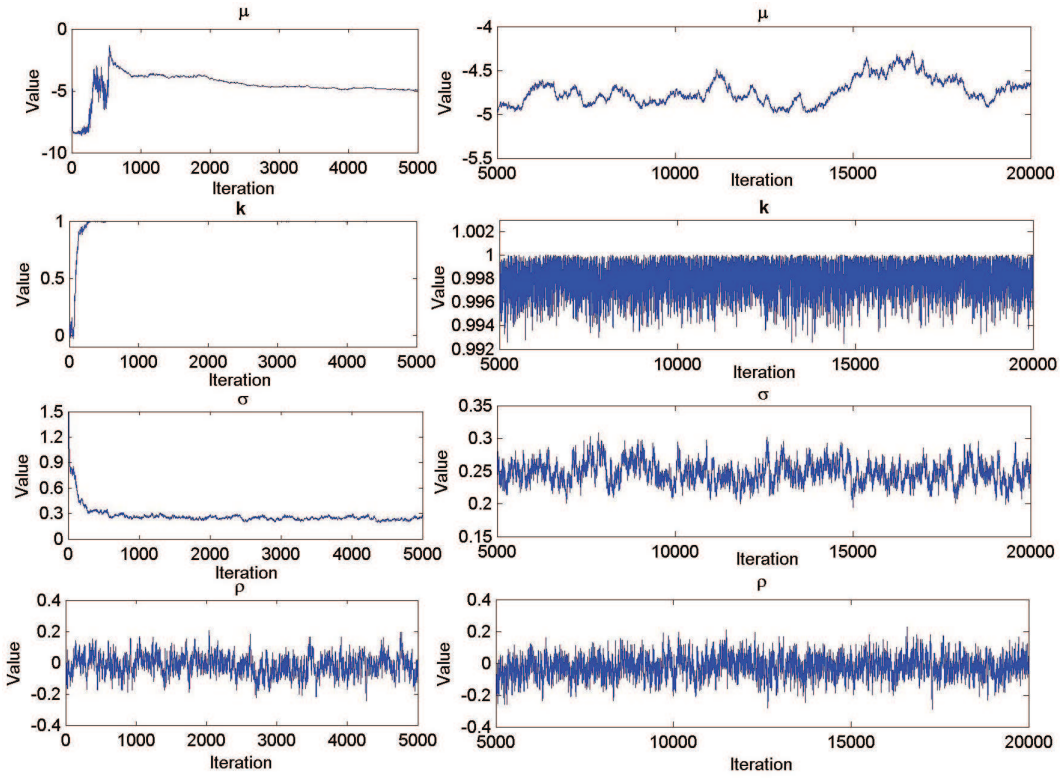

Figure 4.12: Sample paths of model parameters of the stochastic volatility model using MCMC in WinBUGS with leverage effect.

	Mean	Std.	95% Credible Interval	Sample
μ	-4.738	0.1460	(-4.9550,-4.3960)	15,000
k	0.998	0.0013	(0.9951,0.9999)	15,000
σ	0.247	0.01675	(0.2170,0.2822)	15,000
ρ	-0.023	0.0651	(-0.1545,0.1045)	15,000

Table 4.5: Summary statistics of parameter posterior estimates of the stochastic volatility model using MCMC in WinBUGS with leverage effect.

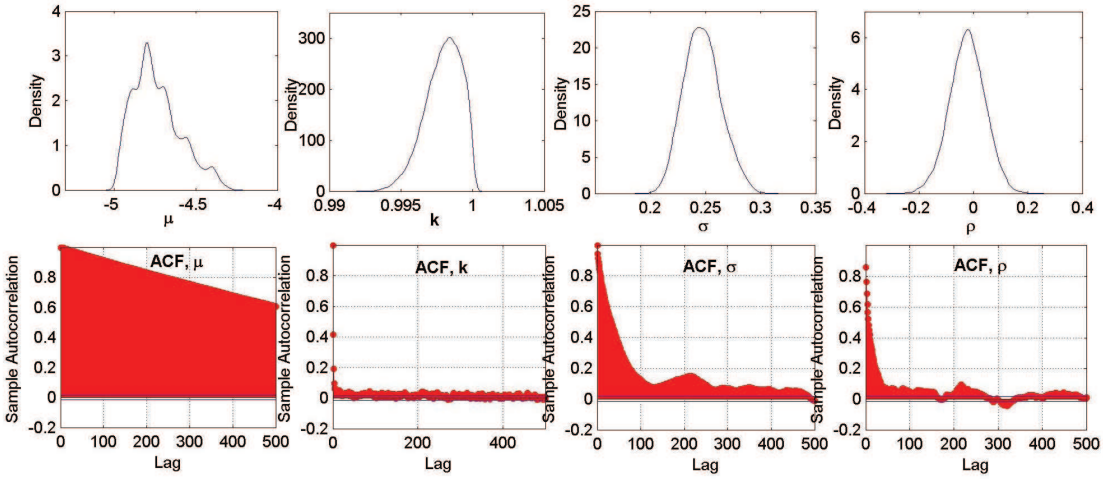

Figure 4.13: Posterior densities (top) and autocorrelation functions (bottom) of model parameters of the stochastic volatility model using MCMC in WinBUGS with leverage effect.

4.4 MCMC in Matlab

Although WinBUGS is a convenient tool for implementing the MCMC methods, it is a "black-box" that does not allow us to know precisely how the algorithms work. When the resulting estimates make less sense, it is difficult to carry out further in-depth analysis. Most importantly, the number of mathematic functions in WinBUGS is very limited so that some complex models may not be implemented. This will be the case when we estimate the Cairns term structure models (which is necessary to incorporate the numerical integral) using MCMC in later chapters.

In this section, we consider coding the MH algorithm in Matlab (or one can use any other programming language) to re-estimate the mean-reverting AR(1) model in (4.6) with the same priors and dataset. The distinct advantage of using Matlab is the flexibility that to be able to associate the model with a wide range of mathematical functions and algorithms available in Matlab.

4.4.1 Likelihood, Prior and Posterior

Suppose that observed data $X = (X(1), X(2), \ldots, X(N))$ follow the mean-reverting normal AR(1) process such that

$$X(i) = \mu + k(X(i-1) - \mu) + \sigma Z(i), \text{ where } Z(i) \sim i.i.d.\ N(0,1), \tag{4.15}$$

it can be proved that the likelihood of X is

$$L(X|\theta) = L_1(X(1)|\theta) \cdot L_2(X(2), \ldots, X(N)|\theta, X(1)), \tag{4.16}$$

where

$$L_1(X(1)|\theta) = \left(\frac{2\pi\sigma^2}{1-k^2}\right)^{-1/2} \exp\left[-\frac{1-k^2}{2\sigma^2}(X(1)-\mu)^2\right],$$

(4.17)

$$L_2(X(2), \ldots, X(N)|\theta) = \prod_{i=2}^{N} (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{1}{2\sigma^2}(X(i)-\mu-k(X(i-1)-\mu))^2\right].$$

Therefore, with the priors specified in (4.7), the full conditional posterior distributions of the model parameters can be written as

$$\begin{aligned} f(\mu|X,k,\tau) &\propto f(X|\mu,k,\tau) \cdot f(\mu) \\ f(k|X,\mu,\tau) &\propto f(X|\mu,k,\tau) \cdot f(k) \\ f(\tau|X,\mu,k) &\propto f(X|\mu,k,\tau) \cdot f(\tau), \end{aligned} \tag{4.18}$$

where $\tau = 1/\sigma^2$. Note that closed-form solutions of (4.18) do not necessarily to be derived since we will use the MH algorithm for the estimation.

4.4.1.1 The Metropolis-Hastings Algorithm

Initially, we choose the normal proposal distributions q_μ, q_k and q_τ to sample candidate points for the parameters μ, k and τ respectively for the next step $t+1$ as follows.

$$\begin{aligned} q_\mu(\cdot|\mu_t) &\equiv N(\mu_t, vol_\mu^2), \\ q_k(\cdot|k_t) &\equiv N(k_t, vol_k^2), \\ q_\tau(\cdot|\tau_t) &\equiv N(\tau_t, vol_\tau^2), \end{aligned} \tag{4.19}$$

where vol_μ, vol_k and vol_τ are some constant values which will be set as $0.05, 0.05$ and 0.5 for our estimation in order for the acceptance rate to be in an appropriate range.

For the initial values, we define

$$\mu_0 = -2.0, \quad k_0 = 0.5, \quad \tau_0 = 1/0.4^2. \tag{4.20}$$

Finally, the MH algorithm is outlined below.

1. Set $t = 0$ and initialise the chain to $\theta_0 = (\mu_0, k_0, \tau_0)$.

2. Generate a candidate point μ^* from the proposal distribution $q_\mu(\cdot|\mu_t)$ in (4.19). Then,

   - Generate U from $\sim U(0, 1)$.
   - Compute the acceptance probability

$$\begin{aligned} \alpha_1(\mu_t, \mu^*) &= \min\left\{1, \frac{f(\mu^*|X, k_t, \tau_t) \cdot q_\mu(\mu_t|\mu^*)}{f(\mu_t|X, k_t, \tau_t) \cdot q_\mu(\mu^*|\mu_t)}\right\} \\ &= \min\left\{1, \frac{f(\mu^*|X, k_t, \tau_t)}{f(\mu_t|X, k_t, \tau_t)}\right\} \end{aligned}$$

     due to the symmetry of the proposal distribution.

   - If $U \leq \alpha_1(\mu_t, \mu^*)$, then $\mu_{t+1} = \mu^*$; otherwise $\mu_{t+1} = \mu_t$.

3. Generate a candidate point k^* from the proposal distribution $q_k(\cdot|k_t)$ in (4.19). Then,

   - Generate U from $\sim U(0, 1)$.
   - Compute the acceptance probability

$$\begin{aligned} \alpha_2(k_t, k^*) &= \min\left\{1, \frac{f(k^*|X, \mu_{t+1}, \tau_t) \cdot q_k(k_t|k^*)}{f(k_t|X, \mu_{t+1}, \tau_t) \cdot q_k(k^*|k_t)}\right\} \\ &= \min\left\{1, \frac{f(k^*|X, \mu_{t+1}, \tau_t)}{f(k_t|X, \mu_{t+1}, \tau_t)}\right\} \end{aligned}$$

     due to the symmetry of the proposal distribution.

   - If $U \leq \alpha_2(k_t, k^*)$, then $k_{t+1} = k^*$; otherwise $k_{t+1} = k_t$.

4. Generate a candidate point τ^* from the proposal distribution $q_\tau(\cdot|\tau_t)$ in (4.19). Then,

- Generate U from $\sim U = (0, 1)$.
- Compute the acceptance probability

$$\begin{aligned} \alpha_3(\tau_t, \tau^*) &= \min\left\{1, \frac{f(\tau^*|X, \mu_{t+1}, k_{t+1}) \cdot q_\tau(\tau_t|\tau^*)}{f(\tau_t|X, \mu_{t+1}, k_{t+1}) \cdot q_\tau(\tau^*|\tau_t)}\right\} \\ &= \min\left\{1, \frac{f(\tau^*|X, \mu_{t+1}, k_{t+1})}{f(\tau_t|X, \mu_{t+1}, k_{t+1})}\right\} \end{aligned}$$

  due to the symmetry of the proposal distribution.
- If $U \leq \alpha_3(\tau_t, \tau^*)$, then $\tau_{t+1} = \tau^*$; otherwise $\tau_{t+1} = \tau_t$.

5. Set $t = t + 1$ and repeat step 2 through 5 until convergence.

4.4.1.2 Results

Figures 4.14 and 4.15 and Table 4.6 show all the resulting estimates of the mean-reverting AR(1) model using the MH algorithm in Matlab (the acceptance rates are controlled at around 40% to 50%). Comparing to the previous estimates in WinBUGS (Figures 4.3 and 4.4 and Table 4.2), we can see that they are generally very similar except for the high autocorrelation functions for the first 10 lags of the estimation in Matlab.

In summary, Chapter 4 presented MCMC which is the core methodology that will be used for fitting the two-factor Cairns term structure model in subsequent chapters. A number of examples with simple models were illustrated in order to provide a ground and familiarity with the methodology before moving to consider estimation with more complex models.

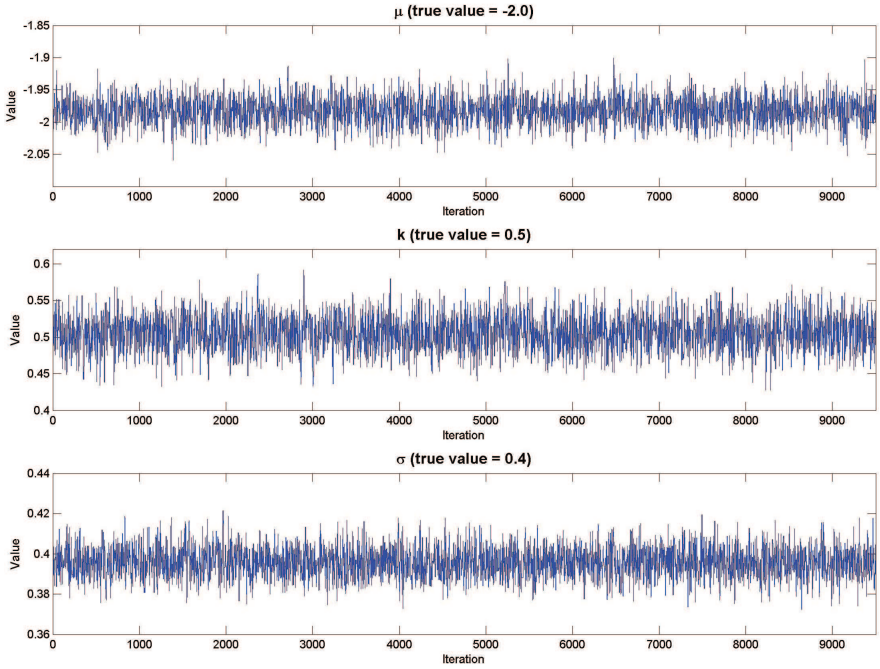

Figure 4.14: Sample paths of model parameters of the mean reverting AR(1) model using the MH algorithm in Matlab.

	(True value)	Mean	Std.	95% Credible Interval	Acceptance rate	Sample
μ	(-2.0)	-1.983	0.0203	(-2.0226,-1.9429)	42.74%	9,500
k	(0.5)	0.506	0.02333	(0.4607,0.5516)	46.85%	9,500
σ	(0.4)	0.396	0.0072	(0.3820,0.4106)	47.61%	9,500

Table 4.6: Summary statistics of parameter posterior estimates of the mean reverting AR(1) model using the MH algorithm in Matlab.

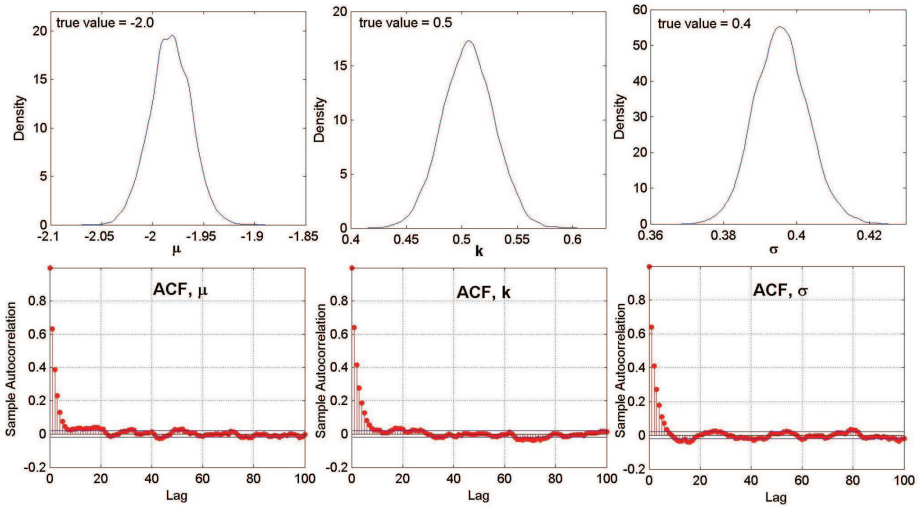

Figure 4.15: Posterior densities and autocorrelation functions of model parameters of the mean reverting AR(1) model using the MH algorithm in Matlab.

Chapter 5

Estimating Two-Factor Cairns Term Structure Models using Markov Chain Monte Carlo: Simulated Data

In this chapter, we estimate the time-varying latent variables and model parameters of the two-factor Cairns term structure models (Cairns, 2004) using Markov Chain Monte Carlo (MCMC). The model is an arbitrage-free model developed under the positive-interest rate framework (Flesaker and Hughston, 1996) for use in long-term risk management. To begin with, we establish an estimation framework by assuming that bond prices from the Cairns model are conditionally normally distributed. Then, we prepare simulated data where the latent state variables are simulated from the exact solution of a two-dimensional Ornstein-Uhlenbeck process. The simulated data are used in the first place in order to validate the simulation algorithm and to ensure that it can result in reasonable and reliable estimates before estimating the model using real market data in next chapter.

Next, theoretical bond prices are numerically computed using the Trapezoidal rule since it is most convenient for the programming and, more importantly, we found that by this simple method there is no significant difference to the prices compared to using more complicated techniques such as the adaptive Simpson quadrature (See Appendix C). Consequently, the full joint posterior density of the Cairns bond price is

derived and hence the latent variables and model parameters are estimated where the two main MCMC methods play a key role for our estimation. Initially, we attempt to employ the Gibbs sampler in which it can be used when a sample can be drawn easily from the conditional posterior. Unfortunately, sampling from the posterior is not straightforward so that, to be able to implement Gibbs sampler, we approximate the posterior using a quadratic form and apply the finite difference method for computing the corresponding partial derivatives. Apparently, we observed that if such approximations are not accurate enough, the resulting estimates might be biased or even go wrong although fast convergence can be achieved. Next, we therefore estimate the full model using the Metropolis-Hastings algorithm which is facilitated by a proposal distribution. The algorithm enables us to avoid drawing a sample from the full conditional posterior but we are required to calibrate variances of the proposal distribution, which is not trivial due to high complexity of the interactions among all the model parameters and latent variables.

The results by the standard Metropolis-Hastings are fairly acceptable but it is evident that the chains converge rather slowly. Accordingly, we also improve the chain convergence by repameterising the bond posterior and using the adaptive proposal distribution with a blocking strategy.

5.1 Framework

Suppose that $(\Omega, \mathcal{F}, P)$ is a probability space. We first set up an estimation framework by assuming interest rates in the market follow the Cairns model such that the observations

$$P(t, \tau_{tj}) = C(\tau_{tj}; X(t), \theta) + \varepsilon(t, j), \tag{5.1}$$

where $\Theta = (\mathbf{X}, \theta)$ with $\theta = (\beta, \alpha_1, \alpha_2, \sigma_1, \sigma_2, \rho, \gamma_1, \gamma_2)'$ being the model parameter vector and $\mathbf{X} = \{X(1), \ldots, X(M)\}$, $X(t) = (X_1(t), X_2(t))'$ being the latent variables which follow $dX_i(t) = \alpha_i(\gamma_i - X_i(t))dt + \sum_{j=1}^{2} \sigma_{ij} dW_j(t)$, where $W_1(t)$ and $W_2(t)$ are two independent Wiener processes with respect to a filtration $(\mathcal{F}_t)_{t\geq 0}$ under the real world probability P. Also, $\varepsilon(t, j) \sim$ i.i.d. $N(0, \sigma_\varepsilon^2)$, $P(t, \tau_{tj})$ is a zero-coupon bond price at time t for a bond that pays 1 at time $t + \tau_{tj}$ of the maturities τ_{tj}, for

$j = 1, 2, \ldots, N_t$, $C(\tau_{tj}; X(t), \theta)$ is the bond price by the two-factor Cairns model, i.e.

$$C(\tau, x, \theta) = \frac{\int_\tau^\infty H(u, x)du}{\int_0^\infty H(u, x)du}, \tag{5.2}$$

where

$$H(u, x) = \exp\Big[- \beta u + \sum_{i=1}^{2} \sigma_i x_i \, e^{-\alpha_i u} - \frac{1}{2} \sum_{i,j=1}^{2} \frac{\rho_{ij}\sigma_i\sigma_j}{\alpha_i + \alpha_j} e^{-(\alpha_i+\alpha_j)u} \Big]. \tag{5.3}$$

With this setting, we have $P(t, \tau_{tj}) \sim N(C(\tau_{tj}; X(t), \theta), \sigma_\varepsilon^2)$.

5.2 Simulated Data

In the early stage, we consider the estimation of the latent variables and model parameters with a simulated dataset which will allow us to check the accuracy of the algorithm with reference to the true values for $\mathbf{X}$ and θ. To simulate the bond prices according to (5.1), we initially define 20 constant maturities ($N_t = 20$, for all t): 0.25, 0.5, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 12.5, 15.0, 17.5, 20.0, 22.5, 25.0, 27.5, 30.0 years, and then generate unit normal random variables $\varepsilon(t, j)$, for $j = 1, 2, \ldots, 20$ with the mean and covariance matrix:

$$\mu = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_\varepsilon & 0 & \cdots & 0 \\ 0 & \sigma_\varepsilon & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_\varepsilon \end{bmatrix}, \text{ where } \sigma_\varepsilon = 0.001.$$

Clearly, we assume here that the bond prices of each time t are independent with a fixed normal randomness. Additionally, the other model parameter values are chosen with respect to Cairns (2004) as

$$\beta = 0.04, \alpha_1 = 0.6, \alpha_2 = 0.06, \sigma_1 = 0.6, \sigma_2 = 0.4, \rho = -0.5,$$

whereas the latent variables will be simulated from the exact solution of a two-dimensional Ornstein-Uhlenbeck process. With these values, simulated monthly bond

prices for 1,000 months can be achieved.

5.2.1 Simulated Latent Variables

Instead of using the Euler discretisation, we know that the latent variables $(X_1(t), X_2(t))'$ follow the two-dimensional Ornstein-Uhlenbeck process so that they can be simulated from the exact solution. Thus, we first introduce the following proposition.

Proposition 5.1. *[The Exact Solution of a Two-Dimensional Ornstein-Uhlenbeck Process] Suppose that $X(t) = (X_1(t), X_2(t))'$ follows a two-dimensional Ornstein-Uhlenbeck process such that*

$$d\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix} \begin{pmatrix} \gamma_1 - X_1(t) \\ \gamma_2 - X_2(t) \end{pmatrix} dt + \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} d\begin{pmatrix} W_1(t) \\ W_2(t) \end{pmatrix}, \text{ or}$$

$$dX_i(t) = \alpha_i(\gamma_i - X_i(t))dt + \sum_{j=1}^{2} \sigma_{ij} dW_j(t),$$

where $X_1(0) = \hat{x}_1, X_2(0) = \hat{x}_2$, $W_1(t)$ and $W_2(t)$ are two independent Wiener processes. Then, the exact solution of $X(t)$ can be found and achieved at

$$X_i(t) = \gamma_i + (X_i(0) - \gamma_i)e^{-\alpha_i t} + \sum_{j=1}^{2} \sigma_{ij} \int_0^t e^{-\alpha_i(t-s)} dW_j(s). \tag{5.4}$$

Hence, $(X_1(t), X_2(t))'$ is bivariate normal with

$$\begin{aligned} \mathbb{E}(X_i(t)) &= \gamma_i + (X_i(0) - \gamma_i)e^{-\alpha_i t}, \\ Var(X_i(t)) &= \frac{(\sigma_{i1}^2 + \sigma_{i2}^2)}{2\alpha_i}(1 - e^{-2\alpha_i t}), \\ Cov(X_1(t), X_2(t)) &= \frac{\sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22}}{\alpha_1 + \alpha_2}(1 - e^{-(\alpha_1+\alpha_2)t}). \end{aligned}$$

Proof. Let $Y_i(t) = e^{\alpha_i t}(X_i(t) - \gamma_i)$. Therefore,

$$dY_i(t) = \alpha_i Y_i(t)dt + e^{\alpha_i t} dX_i(t)$$

$$
\begin{aligned}
&= \alpha_i Y_i(t)dt + e^{\alpha_i t}\left(\alpha_i(\gamma_i - X_i(t))dt + \sum_{j=1}^{2}\sigma_{ij}dW_j(t)\right)\\
&= e^{\alpha_i t}\sum_{j=1}^{2}\sigma_{ij}dW_j(t).
\end{aligned}
$$

By Itô formula, we get

$$
\begin{aligned}
Y_i(t) &= Y_i(0) + \sum_{j=1}^{2}\sigma_{ij}\int_0^t e^{\alpha_i s}dW_j(s)\\
e^{\alpha_i t}(X_i(t)-\gamma_i) &= (X_i(0)-\gamma_i) + \sum_{j=1}^{2}\sigma_{ij}\int_0^t e^{\alpha_i s}dW_j(s)\\
X_i(t) &= \gamma_i + (X_i(0)-\gamma_i)e^{-\alpha_i t} + \sum_{j=1}^{2}\sigma_{ij}\int_0^t e^{-\alpha_i(t-s)}dW_j(s).
\end{aligned}
$$

Clearly, $\mathbb{E}(X_i(t)) = \gamma_i + (X_i(0)-\gamma_i)e^{-\alpha_i t}$ since $\int_0^t e^{-\alpha_i(t-s)}dW_j(s) \sim N(0, \int_0^t e^{-2\alpha_i(t-s)}ds)$ by the Itô isometry. Next, we define $D_i = \sum_{j=1}^{2}\sigma_{ij}\int_0^t e^{-\alpha_i(t-s)}dW_j(s) = \sigma_{i1}d_{i1} + \sigma_{i2}d_{i2}$, where $d_{ij} = \int_0^t e^{-\alpha_i(t-s)}dW_j(s)$. So,

$$
\begin{aligned}
\mathbb{E}(d_{ij}^2) &= \int_0^t e^{-2\alpha_i(t-s)}ds = \frac{1}{2\alpha_i}(1-e^{-2\alpha_i t})\\
\mathbb{E}(d_{i1}d_{i2}) &= 0 \ \text{ (since } d_{i1}, d_{i2} \text{ are independent)},
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{E}(d_{11}d_{21}) &= \mathbb{E}\left(\int_0^t e^{-\alpha_1(t-s)}dW_1(s)\int_0^t e^{-\alpha_2(t-s)}dW_1(s)\right)\\
&= \int_0^t e^{-(\alpha_1+\alpha_2)(t-s)}ds\\
&= \frac{1}{\alpha_1+\alpha_2}(1-e^{-(\alpha_1+\alpha_2)t}) = \mathbb{E}(d_{12}d_{22}).
\end{aligned}
$$

Accordingly,

$$
Var(X_i(t)) = Var(D_i) = \sigma_{i1}^2\mathbb{E}(d_{i1}^2) + \sigma_{i2}^2\mathbb{E}(d_{i2}^2)
$$

$$
Var(D_1) = \frac{(\sigma_{11}^2+\sigma_{12}^2)}{2\alpha_1}(1-e^{-2\alpha_1 t})
$$

$$Var(D_2) = \frac{(\sigma_{21}^2 + \sigma_{22}^2)}{2\alpha_2}(1 - e^{-2\alpha_2 t}),$$

$$\begin{aligned} Cov(X_1(t), X_2(t)) &= Cov(D_1, D_2) = \mathbb{E}(D_1 D_2) - \mathbb{E}(D_1)\mathbb{E}(D_2) = \mathbb{E}(D_1 D_2) \\ &= \sigma_{11}\sigma_{21}\mathbb{E}(d_{11}d_{21}) + \sigma_{12}\sigma_{22}\mathbb{E}(d_{12}d_{22}) \\ &= \frac{\sigma_{11}\sigma_{21} + \sigma_{12}\sigma_{22}}{\alpha_1 + \alpha_2}(1 - e^{-(\alpha_1+\alpha_2)t}). \end{aligned}$$

□

The latent variables $X(t)$ under the real world probability measure P can be simulated from the exact solution in Proposition 5.1. Given $\gamma_1 = \gamma_2 = 0$ and the instantaneous correlation matrix

$$\rho = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} = AA^T, \quad \text{where} \quad A = \begin{bmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{bmatrix},$$

it follows that

$$\begin{aligned} X_1(t) &= X_1(0)e^{-\alpha_1 t} + \int_0^t e^{-\alpha_1(t-s)} dW_1(s), \\ X_2(t) &= X_2(0)e^{-\alpha_2 t} + \rho \int_o^t e^{-\alpha_2(t-s)} dW_1(s) + \sqrt{1-\rho^2} \int_0^t e^{-\alpha_2(t-s)} dW_2(s). \end{aligned} \tag{5.5}$$

Specifically,

$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} \sim N_2 \left(\begin{pmatrix} X_1(0)e^{-\alpha_1 t} \\ X_2(0)e^{-\alpha_2 t} \end{pmatrix}, \begin{pmatrix} \frac{1}{2\alpha_1}(1 - e^{-2\alpha_1 t}) & \frac{\rho}{\alpha_1+\alpha_2}(1 - e^{-(\alpha_1+\alpha_2)t}) \\ \frac{\rho}{\alpha_1+\alpha_2}(1 - e^{-(\alpha_1+\alpha_2)t}) & \frac{1}{2\alpha_2}(1 - e^{-2\alpha_2 t}) \end{pmatrix} \right).$$

Likewise, from time t_k to t_{k+1}, where $\Delta t = t_{k+1} - t_k$, we will have

$$\begin{pmatrix} X_1(t_{k+1}) \\ X_2(t_{k+1}) \end{pmatrix} \sim N_2 \left(\begin{pmatrix} X_1(t_k)e^{-\alpha_1 \Delta t} \\ X_2(t_k)e^{-\alpha_2 \Delta t} \end{pmatrix}, \Sigma_{\Delta t} \right),$$

where

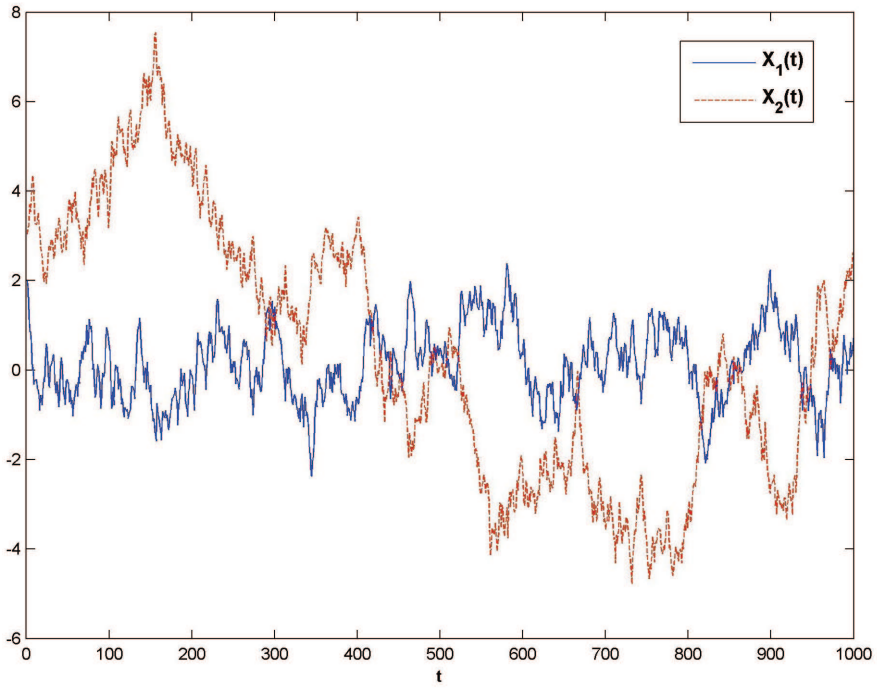

Figure 5.1: The simulated $X_1(t)$(solid) and $X_2(t)$(dotted) from the exact solution for $t = 1, ..., 1000, \Delta t = 1/12$ with $\beta = 0.04, \alpha_1 = 0.6, \alpha_2 = 0.06, \sigma_1 = 0.6, \sigma_2 = 0.4, \rho = -0.5, \gamma_1 = 0$ and $\gamma_2 = 0$.

$$\Sigma_{\Delta t} = \begin{pmatrix} \frac{1}{2\alpha_1}(1 - e^{-2\alpha_1 \Delta t}) & \frac{\rho}{\alpha_1+\alpha_2}(1 - e^{-(\alpha_1+\alpha_2)\Delta t}) \\ \frac{\rho}{\alpha_1+\alpha_2}(1 - e^{-(\alpha_1+\alpha_2)\Delta t}) & \frac{1}{2\alpha_2}(1 - e^{-2\alpha_2 \Delta t}) \end{pmatrix}.$$

Figure 5.1 shows the simulation results of $X_1(t)$ and $X_2(t)$ from $t = 1, 2, ..., M = 1,000$ with time step $\Delta t = 1/12$, given the initial and parameter values: $X_1(1) = 2, X_2(1) = 3, \alpha_1 = 0.6, \alpha_2 = 0.06, \rho = -0.5$.

5.3 Numerical Bond Prices

It is obvious that the solution of the Cairns bond price in (5.2) is not available in a closed-form and hence it needs to be approximated. Thus, we consider four numerical integration methods to estimate an integral of the function $H(u, x)$ in the bond price formula: the Trapezoidal rule, Simpson's rule, Boole's rule and adaptive Simpson quadrature (see Appendix C). Although the adaptive Simpson quadrature is an efficient method that does not rely on a step size, it is far from suitable for a routine simulation as it spends considerably more time than the other methods to

evaluate an integral. In terms of accuracy, we found that no method dominates one another since, given the same parameter and latent variable values, the differences of the bond yields of 20 maturities for a single date are less than 1.0 basis points among all the methods. For simplicity of programming, the Trapezoidal rule is accordingly chosen for our simulation.

Assume that $I = \int_{\tau_0}^{\tau_1} f(u)du$ for some continuous and differentiable function $f(u)$, where $u_0 = \tau_0$ and $u_n = \tau_1$ for some positive integer n with $u_k = \tau_0 + (\tau_1 - \tau_0)k/n$ for $k = 1, \ldots, n-1$. Also let $f_k = f(u_k)$ for $k = 0, \ldots, n$. Then, by the Trapezoidal rule, we will have

$$\begin{aligned} I &\approx \sum_{j=1}^{n} \frac{h}{2}(f_j + f_{j-1}) \\ &= h\left(\frac{1}{2}f_0 + f_1 + \ldots + f_{n-1} + \frac{1}{2}f_n\right), \end{aligned} \tag{5.6}$$

where a step size $h = (\tau_1 - \tau_0)/n$.

5.4 Full Joint Posterior Density of the Cairns Bond Price

Initially, we will look at the mean-reverting bivariate vector autoregressive VAR(1) model since it can be thought of as a model for the latent variables in discrete-time version. The derived likelihood will be part of the full posterior distribution of the Cairns bond price which is required for the MCMC simulation.

Proposition 5.2. *[Likelihood of the Bivariate Normal VAR(1) Model] Suppose that* $X(t) = (X_1(t),\ X_2(t))'$ *follows*

(5.7)
$$\begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} k_1 & 0 \\ 0 & k_2 \end{pmatrix} \begin{pmatrix} X_1(t-1) - \mu_1 \\ X_2(t-1) - \mu_2 \end{pmatrix} + \begin{pmatrix} Z_1(t) \\ Z_2(t) \end{pmatrix}, \ \textit{or}$$

$$X(t) = \mu + K(X(t-1) - \mu) + Z(t),$$

where $Z(t) \sim N_2(0, \Sigma)$ *and* $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$ *is a covariance matrix.*

Then, the likelihood of $\boldsymbol{X} = \{X(1), X(2), \ldots, X(M)\}$*, given the parameter vector* $\theta = (\mu, K, \Sigma)$*, is*

$$\begin{aligned} f(\boldsymbol{X}|\mu, K, \Sigma) &= L_1(X(1)|\theta) \cdot L_2(X(2), \ldots, X(M)|\theta, X(1)) \\ &= (2\pi)^{-1}|\Omega_x|^{-1/2} \exp\left\{-\frac{1}{2}(X(1)-\mu)'\Omega_x^{-1}(X(1)-\mu)\right\} \cdot \\ &\quad (2\pi)^{-(M-1)}|\Sigma|^{-\frac{1}{2}(M-1)} \exp\left\{-\frac{1}{2}\sum_{t=2}^{M}\hat{Z}(t)'\Sigma^{-1}\hat{Z}(t)\right\}, \end{aligned} \tag{5.8}$$

where $\hat{Z}(t) = X(t) - \mu - K(X(t-1) - \mu)$ *and*

$$\Omega_x = \begin{pmatrix} \frac{\sigma_{11}}{1-k_1^2} & \frac{\sigma_{12}}{1-k_1k_2} \\ \frac{\sigma_{12}}{1-k_1k_2} & \frac{\sigma_{22}}{1-k_2^2} \end{pmatrix}.$$

Proof. The likelihood of a model of this kind typically consists of two parts which are the unconditional and conditional likelihoods. The conditional $L_2(\cdot)$ is obvious from the model setting in (5.7). For $L_1(\cdot)$, the unconditional likelihood of $X(1)$, we have

$$\mathbb{E}(X(t)) = \mu + K\mathbb{E}(X(t-1) - \mu),$$

and by expanding $X(t-1)$ in terms of its previous values, we know that

$$X(t) = \mu + \sum_{j=0}^{\infty} K^{(j)} Z(t-j).$$

Hence,

$$\begin{aligned} \Omega_x = Var(X(t)) &= \sum_{j=0}^{\infty} K^{(j)} \Sigma K'^{(j)} \\ &= \sum_{j=0}^{\infty} \begin{pmatrix} (k_1^2)^j\sigma_{11} & (k_1k_2)^j\sigma_{12} \\ (k_1k_2)^j\sigma_{21} & (k_2^2)^j\sigma_{22} \end{pmatrix} \end{aligned}$$

$$= \begin{pmatrix} \sigma_{11}/(1-k_1^2) & \sigma_{12}/(1-k_1k_2) \\ \sigma_{21}/(1-k_1k_2) & \sigma_{22}/(1-k_2^2) \end{pmatrix}$$

since $\frac{1}{1-x} = 1 + x + x^2 + \ldots$. □

Providing the framework in (5.1), the full joint posterior density of the Cairns bond price can be written as

$$\begin{aligned} f(\Theta|\mathbb{P}) \propto f(\mathbb{P},\Theta) &= f(\mathbb{P}|\Theta)f(\Theta) \\ &= \prod_{t=1}^{M} f_1(\mathbb{P}|X(t),\theta) \\ &\quad \times \prod_{t=2}^{M} f_{2c}(X(t)|X(t-1),\theta_2) \times f_{2u}(X(1)|\theta) \times f_0(\theta), \end{aligned} \tag{5.9}$$

where $\mathbb{P}$ is all bond price data, f_1, f_{2u} and f_{2c} are the normal density functions, f_0 is the prior density function and $\Theta = (\mathbf{X}, \theta)$, where $\mathbf{X} = \{X(1), X(2), ..., X(M)\}, \theta = \theta_1 \cup \theta_2$, where $\theta_1 = \{\beta, \sigma_1, \sigma_2, \sigma_\varepsilon\}, \theta_2 = \{\alpha_1, \alpha_2, \rho, \gamma_1, \gamma_2\}$.

We partition the parameter vector θ into θ_1 and θ_2 to make clear which group of parameters is in the dynamic of the latent variables. In addition, it can be observed that the full posterior distribution (5.9) consists of four main components:

- the likelihood of the pricing data; f_1 (measurement equation),
- the conditional likelihood of the latent variables $X(t)$; f_{2c} (transition equation),
- the unconditional likelihood of $X(1)$; $f_{2u}(X(1)|\theta)$, and
- the prior density of of model parameters $f_0(\theta)$.

Since the likelihood f_{2c} and $f_{2u}(X(1)|\theta)$ are given in Proposition 5.2, the log posterior eventually is

$$F(\Theta) = \log f(\Theta|\mathbb{P}) = k + \log f(\mathbb{P},\Theta), \tag{5.10}$$

where k is a constant and

$$\log f(\mathbb{P},\Theta) = \sum_{t=1}^{M} \log f_1(\mathbb{P}|X(t),\theta)$$

$$+\sum_{t=2}^{M}\log f_{2c}(X(t)|X(t-1),\theta_2)+\log f_{2u}(X(1)|\theta)+\log f_0(\theta)$$

$$\begin{aligned}
&= \sum_{t=1}^{M}\left\{-\frac{N_t}{2}\log(2\pi\sigma_\varepsilon^2)-\frac{1}{2\sigma_\varepsilon^2}\sum_{j=1}^{N_t}\left(P(t,\tau_{tj})-C(\tau_{tj};X(t),\theta)\right)^2\right\}\\
&-(M-1)\log(2\pi)-\frac{(M-1)}{2}\log|\Sigma|-\frac{1}{2}\sum_{t=2}^{M}\hat{Z}(t)'\Sigma^{-1}\hat{Z}(t)\\
&-\log(2\pi)-\frac{1}{2}\log|\Omega_x|-\frac{1}{2}(X(1)-\gamma)'\Omega_x^{-1}(X(1)-\gamma)+\log f_0(\theta)
\end{aligned}$$

$$\begin{aligned}
\log f(\mathbb{P},\Theta) &= -\frac{MN_t}{2}\log(2\pi\sigma_\varepsilon^2)-\frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{M}\sum_{j=1}^{N_t}\left(P(t,\tau_{tj})-C(\tau_{tj};X(t),\theta)\right)^2\\
&-M\log(2\pi)-\frac{1}{2}\log|\Omega_x|-\frac{(M-1)}{2}\log|\Sigma|\\
&-\frac{1}{2}(X(1)-\gamma)'\Omega_x^{-1}(X(1)-\gamma)-\frac{1}{2}\sum_{t=2}^{M}\hat{Z}(t)'\Sigma^{-1}\hat{Z}(t)+\log f_0(\theta),
\end{aligned}$$

where $f_0(\theta)$ is the prior, $C(\tau_{tj};X(t),\theta)$ is theoretical price, $\mathbb{P}$ is all bond price data and $P(t,\tau_{tj})$ is observed price at time t for the maturity τ_{tj} such that

$$P(t,\tau_{tj})\sim N(C(\tau_{tj};X(t),\theta),\sigma_\varepsilon^2),$$

$$\hat{Z}(t)=X(t)-\gamma-K(X(t-1)-\gamma),$$

$$\Sigma=\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}=\begin{pmatrix}\frac{1}{2\alpha_1}(1-e^{-2\alpha_1\Delta t}) & \frac{\rho}{\alpha_1+\alpha_2}(1-e^{-(\alpha_1+\alpha_2)\Delta t})\\ \frac{\rho}{\alpha_1+\alpha_2}(1-e^{-(\alpha_1+\alpha_2)\Delta t}) & \frac{1}{2\alpha_2}(1-e^{-2\alpha_2\Delta t})\end{pmatrix},$$

$$\Omega_x=\begin{pmatrix}\frac{\sigma_{11}}{1-k_1^2} & \frac{\sigma_{12}}{1-k_1k_2}\\ \frac{\sigma_{21}}{1-k_1k_2} & \frac{\sigma_{22}}{1-k_2^2}\end{pmatrix},K=\begin{pmatrix}e^{-\alpha_1\Delta t} & 0\\ 0 & e^{-\alpha_2\Delta t}\end{pmatrix}.$$

5.5 Markov Chain Monte Carlo (MCMC)

The MCMC methods were already described and employed to estimate a number of time-series models in Chapter 4. In this section, we will apply both Metropolis-Hastings (MH) algorithm and Gibbs sampler to the posterior density of the Cairns bond price. Suppose that $\mathbb{P}$ represents all bond price data which are generated from the Cairns model with an unknown parameter and latent variable vector Θ. The MCMC algorithms that can be used for simulating the model parameters and latent variables from the Cairns bond posterior in (5.10) will be discussed as follows.

5.5.1 The Metropolis-Hastings Algorithm

The MH algorithm is a popular updating scheme since it can eliminate the difficulty of drawing a sample from the full posterior distribution (as Gibbs sampler does). Instead of the posterior, a candidate point is first drawn from an arbitrary proposal distribution and then used for calculating the acceptance probability in order to decide a movement of the chain. If the candidate point is rejected, the chain remains the previous value, otherwise it moves to the next state. Ideally, it is often expected to have the proposal distribution as close to the posterior distribution as possible. An appropriate choice of the proposal distributions is therefore important for succeeding in implementing the MH algorithm.

Assume that $\Theta = (\Theta_1, \ldots, \Theta_d)$ is a vector of all unknown quantities in the Cairns term structure model. Then, we know from (5.10) that the posterior distribution of each element of Θ, denoted as Θ_i, is

$$g(\Theta_i|\mathbb{P}, \Theta_{-i}) \quad \propto \quad \exp(\log f(\mathbb{P}, \Theta_i|\Theta_{-i})), \tag{5.11}$$

where Θ_{-i} is a vector of all model parameters and latent variables excluding Θ_i. Remark that the joint density $\log f(\mathbb{P}, \Theta)$ includes $f_0(\theta)$, the prior of model parameters, in which we let

$$f_0(\theta) = f_0(\beta) f_0(\alpha_1) f_0(\alpha_2) f_0(\sigma_1) f_0(\sigma_2) f_0(\rho) f_0(\gamma_1) f_0(\gamma_2)$$

such that

$$\begin{aligned} f_0(\beta), f_0(\alpha_1), f_0(\alpha_2), f_0(\sigma_1), f_0(\sigma_2), &\sim \Gamma(0.01, 100), \\ f_0(\rho) &\sim U[-1, 1], \\ f_0(\gamma_1), f_0(\gamma_2) &\sim N(0, 1.0 \times 10^5). \end{aligned} \tag{5.12}$$

As can be seen, a gamma prior is assigned to the non-negative parameters and a uniform prior with values in the range $[-1, 1]$ to the correlation parameter ρ. Note that σ_z does not appear here since it is a fixed parameter that we do not estimate.

Further, a candidate point $y^*_{\Theta_i}$, for each Θ_i, will be drawn from the normal proposal distribution q_{Θ_i}, with mean depending on its previous value with some constant variance. The detailed procedure of the MH algorithm for the Cairns bond posterior is outlined in Algorithm 1.

Algorithm 1. [Standard Metropolis-Hastings Algorithm]

1. Initialise the chain at $j = 1$ and start the iteration at $j = 2$ and set $i = 1$.
2. Generate a candidate point $y^*_{\Theta_i}$ from the proposal

   $$q_{\Theta_i} \sim N(\Theta_i(j-1), vol^2_{\Theta_i}), \tag{5.13}$$

   where $vol^2_{\Theta_i}$ is some constant volatility.
3. If

   $$\begin{cases} y^*_{\Theta_i} < 0, \text{ for } \Theta_i = \beta, \alpha_1, \alpha_2, \sigma_1, \sigma_2, \\ |y^*_{\Theta_i}| > 1, \text{ for } \Theta_i = \rho, \end{cases}$$

   set $\Theta_i(j) = \Theta_i(j-1)$, and then go to step 5, otherwise go to step 4.
4. Then,
   - Generate U from $\sim U(0, 1)$.
   - Compute the acceptance probability $\eta_{\Theta_i}(\Theta_i(j-1), y^*_{\Theta_i}) =$

     $$\min\left\{1, \frac{g(y^*_{\Theta_i}|\mathbb{P}, \Theta_{-i}) \cdot q_{\Theta_i}(\Theta_i(j-1)|y^*_{\Theta_i})}{g(\Theta_i(j-1)|\mathbb{P}, \Theta_{-i}) \cdot q_{\Theta_i}(y^*_{\Theta_i}|\Theta_i(j-1))}\right\}. \tag{5.14}$$

   - If $U \leq \eta_{\Theta_i}(\Theta_i(j-1), y^*_{\Theta_i})$, then $\Theta_i(j) = y^*_{\Theta_i}$; otherwise $\Theta_i(j) = \Theta_i(j-1)$.

5. Set $i = i + 1$ and repeat step 2 to 4 until $i = d$ (the last element of Θ). For the same iteration j, the recent value of the Θ_i, which is already updated, will be used, rather than its value from the previous step $j - 1$.

6. Set $j = j + 1, i = 1$ and repeat step 2 to 5 until the last iteration (convergence).

Notice that in step 3 we have some knowledge of the parameter values, that is, some are subject to either being non-negative or take values in a certain range. Thus, if the candidate points cannot satisfy these conditions, the chain will not move and immediately go to step 5.

5.5.1.1 Improving the Proposal Distribution

For the normal proposal distribution, it is suggested that the proposal variance should be as close as, or large enough to encompass, the posterior variance. In place of searching for the suitable constant value, two alternatives are considered for improvement as described below.

Algorithm 2. [Hybrid MCMC]

The proposal variance may be approximated from the second derivative of the log posterior density. Specifically, the proposal in (5.13) is replaced by

$$q_{\Theta_i} \sim N\left(\Theta_i(j-1), -\frac{1}{F''_{\Theta_i(j-1)}(\Theta)}\right), \tag{5.15}$$

where $F''_{\Theta_i(j-1)}(\Theta)$ is the second derivative of $\log f(\Theta|\mathbb{P})$ in (5.10) with respect to Θ_i evaluated at $\Theta_i(j-1)$.

Once the normal proposal follows (5.15), the conditional densities of the proposal in (5.14) are no longer symmetric. That is, we now have

$$\begin{aligned} q_{\Theta_i}(\Theta_i(j-1)|y^*_{\Theta_i}) &\sim N\left(y^*_{\Theta_i}, -\frac{1}{F''_{y^*_{\Theta_i}}(\Theta)}\right), \\ q_{\Theta_i}(y^*_{\Theta_i}|\Theta_i(j-1)) &\sim N\left(\Theta_i(j-1), -\frac{1}{F''_{\Theta_i(j-1)}(\Theta)}\right). \end{aligned}$$

More detail regarding Hybrid MCMC can be found in Garmerman and Lopes (2006).

Algorithm 3. [Adaptive Metropolis-Hastings Algorithm]

Referring to Haario et al. (2005), another idea for improving the proposal distribution is to use the empirical variance computed from its previous sample path (recent n_1 values), after using a constant variance for some initial iteration n_0. More precisely, the proposal variance is given by

$$vol^2_{\Theta_i}(j) = \begin{cases} Var_{n_0} & j \leq n_0 \\ k \cdot Var(\Theta_i(j-n_1), \ldots, \Theta_i(j-1)) & j > n_0, \end{cases} \tag{5.16}$$

where Var_{n_0} is some constant value, $Var(\cdot)$ is the sample variance of values in the argument and k is a scaling number. For updating several parameters, the covariance matrix can be used in order to draw a set of candidate points with correlation, i.e.

$$Cov_{\Theta_i}(j) = \begin{cases} Cov_{n_0} & j \leq n_0 \\ k \cdot Cov(\Theta_i(j-n_1), \ldots, \Theta_i(j-1)) & j > n_0, \end{cases} \tag{5.17}$$

where Cov_{n_0} is an initial covariance matrix, $Cov(\cdot)$ is the sample covariance of a series of values in the argument and k is a scaling number.

5.5.2 The Gibbs Sampler

The Gibbs sampler is a special case of the MH algorithm in which the acceptance probability is always equal to 1. In many circumstances, Gibbs sampler can provide much faster convergence than the MH algorithm but the full conditional posterior distribution of each parameter must have been known and can be drawn a sample from. Unfortunately, it is clear that the full log posterior distribution of the Cairns bond price in (5.10) is too complicated to draw a sample. However, it is possible if we approximate the posterior using a quadratic approximation which is in the same form as of a normal distribution. In effect, we also require to compute the derivatives of the posterior, which can be obtained using the finite difference method. Although Gibbs sampler is not the main algorithm used to estimate the Cairns model, it is still worth describing in some detail as follows.

5.5.2.1 Gibbs Sampler for the Latent Variables

The latent variables $X(t) = (X_1(t), X_2(t))'$, for $t = 1, 2, ..., M$, can be sampled as a pair for each t. As aforementioned, to employ Gibbs sampler, we first approximate the bond price log posterior in the quadratic form.

Estimating the log posterior distribution using quadratic approximation

First of all, we define the score function $S(X(t)|\Theta_{-X(t)})$ as

$$S(X(t)|\Theta_{-X(t)}) \equiv \left\{ \frac{\partial F(\Theta)}{\partial X_1(t)}, \frac{\partial F(\Theta)}{\partial X_2(t)} \right\}, \quad \text{for } t = 1, ..., M,$$

where $F(\Theta)$ is the log posterior density according to (5.10) and $\Theta_{-X(t)}$ is all unknown quantities in the model excluding the latent variables being evaluated. At the current evaluating point $\hat{X}(t)$, we know that the second derivative of the log-likelihood is negative definite at the maximum. Therefore, the curvature at $\hat{X}(t)$ can be defined as the Fisher information matrix $I(\hat{X}(t))$ such that

$$I(\hat{X}(t)) \equiv -H(\hat{X}(t)) = \left(\begin{array}{cc} -\frac{\partial^2 F(\Theta)}{\partial X_1(t)^2} & -\frac{\partial^2 F(\Theta)}{\partial X_1(t) \partial X_2(t)} \\ -\frac{\partial^2 F(\Theta)}{\partial X_2(t) \partial X_1(t)} & -\frac{\partial^2 F(\Theta)}{\partial X_2(t)^2} \end{array} \right) \Bigg|_{\hat{X}_1(t), \hat{X}_2(t)},$$

for $t = 1, ..., M$.

By Taylor's expansion to the second-order of $\log f(X(t)|\mathbb{P}, \Theta_{-X(t)})$ around $\hat{X}(t)$, we obtain

$$\begin{aligned} \log f(X(t)|\mathbb{P}, \Theta_{-X(t)}) &\approx \log f(\hat{X}(t)|\mathbb{P}, \Theta_{-X(t)}) \\ &\quad + S(\hat{X}(t)|\Theta_{-X(t)}) \cdot (X(t) - \hat{X}(t)) \\ &\quad - \frac{1}{2}(X(t) - \hat{X}(t))' I(\hat{X}(t))(X(t) - \hat{X}(t)) \\ &\approx \log f(\hat{X}(t)|\mathbb{P}, \Theta_{-X(t)}) \\ &\quad - \frac{1}{2}(X(t) - \hat{X}(t))' I(\hat{X}(t))(X(t) - \hat{X}(t)) \\ \log \frac{f(X(t)|\mathbb{P}, \Theta_{-X(t)})}{f(\hat{X}(t)|\mathbb{P}, \Theta_{-X(t)})} &= -\frac{1}{2}(X(t) - \hat{X}(t))' I(\hat{X}(t))(X(t) - \hat{X}(t)). \end{aligned}$$

Hence,

$$(5.18)\quad \begin{pmatrix} X_1(t) \\ X_2(t) \end{pmatrix} \Big| \Theta_{-X_1(t),-X_2(t)} \sim N_2 \left(\begin{pmatrix} \hat{X}_1(t) \\ \hat{X}_2(t) \end{pmatrix}, I^{-1}(\hat{X}_1(t), \hat{X}_2(t)) \right).$$

for $t = 1, ..., M$.

Approximating the Hessian matrix by finite difference method

It is obvious that the derivatives of $F(\Theta)$ cannot be achieved analytically, but possibly numerically. Hence, the central differences scheme of the finite difference method is taken into an account for this purpose. Given any function $F(X_1, X_2)$ with an evaluating point $x_0 = (\hat{X}_1, \hat{X}_2)'$, the first and second partial derivatives of $F(X_1, X_2)$ can be approximated as

$$\frac{\partial F(x_0)}{\partial X_1} \approx \frac{F(x_0 + (h, 0)') - F(x_0 + (-h, 0)')}{2h}$$

$$\frac{\partial^2 F(x_0)}{\partial X_1^2} \approx \frac{F(x_0 + (h, 0)') - 2F(x_0) + F(x_0 + (-h, 0)')}{h^2}$$

$$\frac{\partial^2 F(x_0)}{\partial X_1 \partial X_2} \approx \frac{F(x_0 + (h, k)') - F(x_0 + (h, -k)') - F(x_0 + (-h, k)') + F(x_0 + (-h, -k)')}{4hk},$$

where the grid points along X_1 and X_2 are equidistant by h and k respectively.

To implement the scheme above, we divide the log posterior distribution in (5.10) into three parts such that

$$F(\Theta) = \log f(\Theta|\mathbb{P}) = k + \log f_1 + \log f_2 + \log f_0,$$

where k is a constant, $\log f_1$ is the log-likelihood of the pricing data, $\log f_2$ includes $\log f_{2c}$ and $\log f_{2u}$ (the conditional and unconditional log-likelihood of the latent variables), and $\log f_0$ is the log prior density of model parameters. Clearly, $\log f_0$ is irrelevant when approximating derivatives of the latent variables since it is treated as a constant for all $F(\Theta)$ evaluated. The two remaining parts will be considered in detail.

$\underline{\log f_1}$

Recall that

$$\log f_1 = \sum_{t=1}^{M} \log f_1(t),$$

$$\log f_1(t) = -\frac{N}{2}\log(2\pi\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2}\sum_{j=1}^{N}\left(P(t,\tau_{tj}) - C(\tau_{tj}; X(t), \theta)\right)^2.$$

As can be observed, $\log f_1$ is also divided into M components. Each depends on the model parameters and the latent variables of only the time t considered. As a consequence, when evaluating the gradient of $F(\Theta)$ with respect to $X(t)$ at any particular time t, $\log f_1(t)$ of other time, which has not been evaluated, will not necessarily be computed. For instance, if we were approximating the first derivative of $F(\Theta)$ with respect to $X_1(1)$, we indeed need to calculate only $\log f_1(1)$ at two evaluating points: $(X_1(1)+h, X_2(1))$ and $(X_1(1)-h, X_2(1))$ since $\log f_1(2), ..., \log f_1(M)$ can be cancelled out when calculating the derivatives with respect to $X_1(1)$. By this way, we are able to considerably reduce the runtime of computing the Hessian matrix because we do not require to numerically approximate all the bond prices $C(\tau_{tj}; X(t), \theta)$ for all t of a whole dataset every time the derivatives with respect to each $X(t)$ being evaluated.

$\underline{\log f_2}$

Recall that

$$\begin{aligned}\log f_2 \;=\; & -M\log(2\pi) - \frac{1}{2}\log|\Omega_x| - \frac{(M-1)}{2}\log|\Sigma| - \frac{1}{2}(X(1)-\gamma)'\Omega_x^{-1}(X(1)-\gamma) \\ & -\frac{1}{2}\sum_{t=2}^{M}\hat{Z}(t)'\Sigma^{-1}\hat{Z}(t),\end{aligned}$$

where $\hat{Z}(t) = X(t) - \gamma - K(X(t-1)-\gamma)$.

Similar to $\log f_1$, the remaining terms in $\log f_2$ which will involve in approximating the derivatives of $F(\Theta)$ with respect to $X(t)$, for $t = 1, ..., M$, will eventually only be

(5.19)

$$\begin{cases} -\frac{1}{2}(X(t)-\gamma)'\Omega_x^{-1}(X(t)-\gamma) - \frac{1}{2}\hat{Z}(t+1)'\Sigma^{-1}\hat{Z}(t+1) & \text{for } t=1, \\ -\frac{1}{2}\hat{Z}(t+1)'\Sigma^{-1}\hat{Z}(t+1) - \frac{1}{2}\hat{Z}(t)'\Sigma^{-1}\hat{Z}(t) & \text{for } t=2,...,M-1, \\ -\frac{1}{2}\hat{Z}(t)'\Sigma^{-1}\hat{Z}(t) & \text{for } t=M. \end{cases}$$

Algorithm 4. [Gibbs Sampler with Quadratic Approximation]

With the quadratic approximation and finite difference method, we are currently able to implement Gibbs sampler to $X(t) = (X_1(t), X_2(t))'$ for $t = 1, \ldots, M$. The algorithm is as follows.

1. Initialise the chains at $j = 1$ and start iterations at $j = 2$ and set $i = 1$.

2. Approximate the Hessian matrix of $F(\Theta)$ with respect to $X(i)$ using the finite difference method and then draw a pair of samples $(X_1(j,i), X_2(j,i))'$.

   - Set up the grid points with equidistance step size h for both $X_1(i)$ and $X_2(i)$:
     - $(X_1(j-1,i), X_2(j-1,i))$,
     - $(X_1(j-1,i) \pm h, X_2(j-1,i)), (X_1(j-1,i), X_2(j-1,i) \pm h)$,
     - $(X_1(j-1,i) \pm h, X_2(j-1,i) \pm h)$.
   - Evaluate $\log f_1$ and $\log f_2$ (as described earlier).
   - Compute the first partial derivative vector $A(j,i)$ and the Hessian matrix $H(j,i)$ with respect to $X(i)$.
   - Compute the adaptive mean

$$\begin{pmatrix} X_1^c(j,i) \\ X_2^c(j,i) \end{pmatrix} = \begin{pmatrix} X_1(j-1,i) \\ X_2(j-1,i) \end{pmatrix} - H^{-1}(j,i)A(j,i).$$

   - Draw a pair of samples from

$$\begin{pmatrix} X_1(j,i) \\ X_2(j,i) \end{pmatrix} \sim N_2\left(\begin{pmatrix} X_1^c(j,i) \\ X_2^c(j,i) \end{pmatrix}, -H^{-1}(j,i)\right).$$

3. Set $i = i + 1$ and repeat step 2 until $i = M$. Note that the most updated sampling values from the previous i will be used for an evaluation of the next i.

4. Set $j = j+1, i = 1$ and repeat step 2 and 3 until the last iteration (convergence).

5.5.2.2 Gibbs Sampler for the Model Parameters

Similar to the latent variables, we also can implement Gibbs sampler to each model parameter θ_i using a quadratic approximation with the log posterior $F(\Theta)$ and the finite difference method to compute the corresponding first and second partial derivatives. Given

$$F'_{\theta_i}(\Theta) = \frac{\partial F(\Theta)}{\partial \theta_i}, F''_{\theta_i}(\Theta) = \frac{\partial^2 F(\Theta)}{\partial \theta_i^2},$$

the θ_i at j-th iteration can be simulated from

$$\theta_i(j) \sim N\left(\theta_i(j-1) - \frac{F'_{\theta_i}(\Theta)}{F''_{\theta_i}(\Theta)}, -\frac{1}{F''_{\theta_i}(\Theta)}\right). \tag{5.20}$$

5.6 Estimating Results on Simulated Data

We are implementing the MCMC methods to the two-factor Cairns term structure model using 100 months of the simulated data (with the latent variables from time $t = 1$ to 100 in Figure 5.1 and model parameter values as specified in Section 5.2.1). With this dataset, 200 latent variables and 8 model parameters are being estimated. Regarding the number of maturities of the bond prices, in fact they can be different for each time t but for convenience we use 20 constant maturities ($N_t = 20$, for all t): 0.25, 0.5, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 12.5, 15.0, 17.5, 20.0, 22.5, 25.0, 27.5, 30.0 years.

The algorithm is coded in Matlab with an interface to a C++ program for the routine of computing the numerical bond prices, which is the most time consuming part of the algorithm. For instance, if we naively coded the MH algorithm to update each model parameter individually for 10,000 iterations with this dataset, we will require to compute the numerical integral 42 million times (= 10,000 iterations $\times$ 1 acceptance probability $\times$ 2 posterior densities $\times$ 100 months $\times$ 21 integrals (for 20 maturities)). As such, designing an efficient algorithm is one crucial step of im-

plementing the MCMC methods to the Cairns bond prices in practice. In our case, the code of this part is accordingly written in C++ which in turn can substantially reduce the computational runtime. Without this, the implementation of MCMC is unlikely to be affordable by entirely coding in Matlab. Moreover, the C++ code is also designed to compute the numerical bond prices for all 20 maturities at a time. By all of these, we are able to run the algorithm over ten times faster than the original pure Matlab code.

The results consist of three parts. First, we fix all the model parameters and estimate only the latent variables using Gibbs sampler with quadratic approximation (Algorithm 4). Second, the full model will be estimated using the standard MH algorithm (Algorithm 1). Finally, we show the results of using the adaptive MH algorithm (Algorithm 3) with reparameterising the bond price log posterior and re-evaluating the prior distributions. Regarding the hybrid MCMC (Algorithm 2), this was also implemented for estimating the model parameters, but satisfying results cannot be achieved and hence they will not be presented here.

5.6.1 Estimating Latent Variables using Gibbs Sampler given Fixed Model Parameters

With quadratic approximation, the latent state variables can be estimated using the Gibbs sampler (Algorithm 4). In this simulation, all the model parameter values are fixed at the true values and we run 5,000 iterations with initial values of $X_1(t) = 2$ and $X_2(t) = 3$ for all $t = 1, \ldots, 100$. $X_1(t)$ and $X_2(t)$ are evaluated as a pair and the Hessian matrix is approximated by the finite difference method with equidistance step size of 0.01.

Figure 5.2 shows 95% credible intervals constructed from the simulated sample paths of all latent variables. As can be seen, the true values reasonably lie in the interval for both $X_1(t)$ and $X_2(t)$. The standard deviations of $X_1(t)$ and $X_2(t)$ are demonstrated in Figure 5.3 and we can notice that all of $X_1(t)$ are higher than the maximum standard deviation of $X_2(t)$ (about 0.0135 at $t = 1$) and change in opposite direction to the true values. Specifically, the standard deviations of $X_1(t)$ tend to increase when its true values decrease and vice versa. Obviously, the standard deviation of $X_2(t)$ is more stable than $X_1(t)$.

In Figure 5.4, the sample paths, posterior densities and autocorrelation functions of $X_1(t)$ and $X_2(t)$ at $t = 20$ are presented. The densities are clearly in a good shape and the autocorrelation functions are substantially low for all lags. Furthermore, we can see that fast convergence can be achieved at the early iterations for both $X_1(20)$ and $X_2(20)$ but not exactly centering at their true values, although they are still within 1.5 standard deviations. One possible reason is that the posterior distribution of the bond price and the corresponding derivatives are both approximated. Therefore, the accuracy of these approximations is essentially important. Due to the high level of non-linearity and complexity of the posterior, the approximations might just be fairly acceptable but not highly accurate.

We conclude from the first result that Gibbs sampler using quadratic approximation with finite difference method performs very well when our log posterior in (5.10) is given fixed model parameters. In all cases, both $X_1(t)$ and $X_2(t)$ converge fast to a level close to the true values just within the first 10 iterations. However, we need to be cautious that the less accurate the derivative approximation is, the more biased the results will be. Hence, it is suggested to use this method only for a burn-in period and subsequently turn to the Metropolis-Hastings algorithm for the remaining iterations.

Apart from this, we in fact also continued to use Algorithm 4 to estimate the model parameters (as described in Section 5.5.2.2) but the resulting estimates (not shown here) are strongly biased, especially for the parameter ρ where we found that it converges to zero which is too far from the true value of -0.5. The reason is simply that the approximation of the derivatives by the finite difference method is inaccurate for ρ.

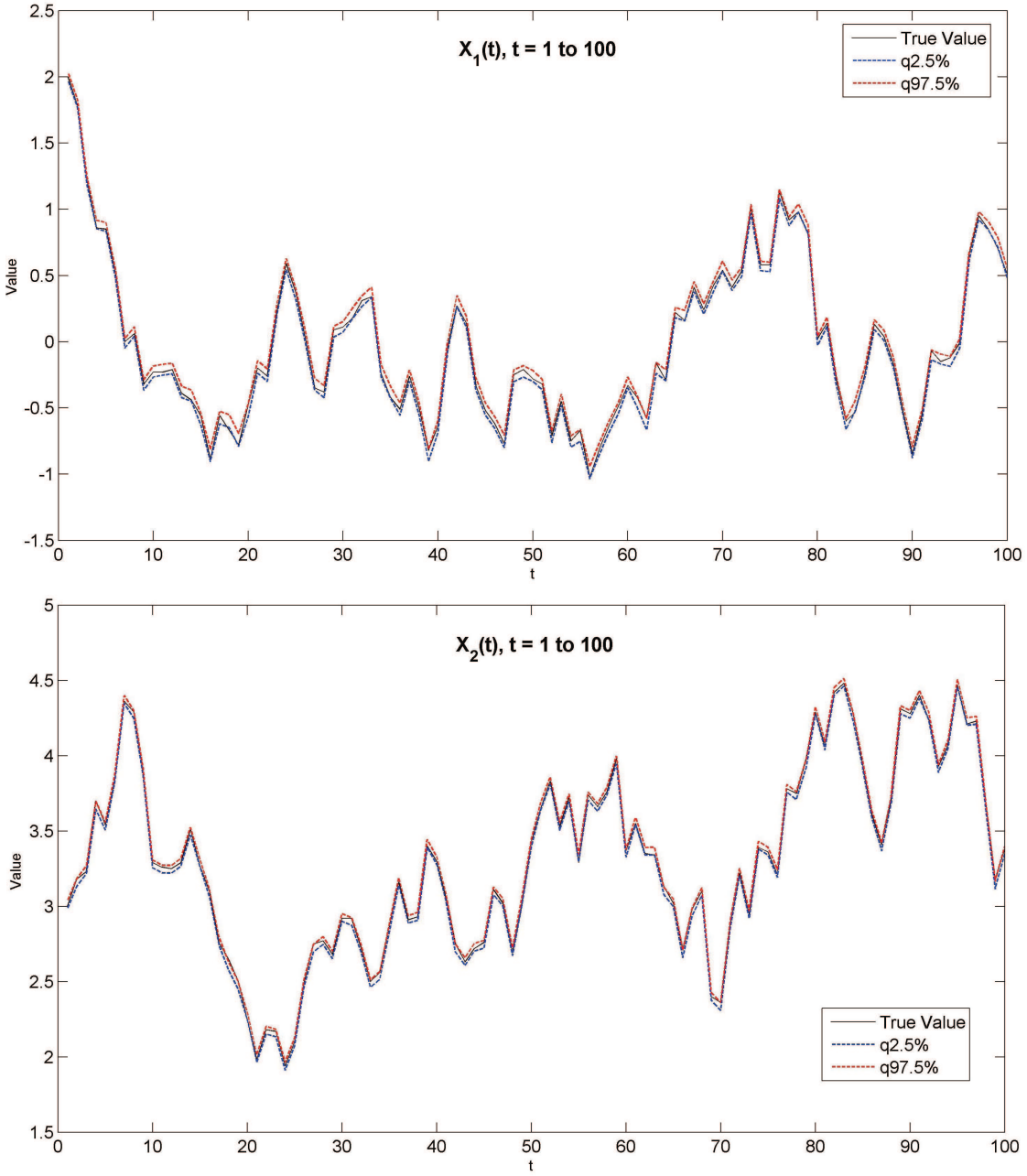

Figure 5.2: Plots of 95% credible interval constructed from the sample paths with the true values of $X_1(t)$(top) and $X_2(t)$(bottom) for $t = 1, \ldots, 100$, (first 100 iterations are excluded) of the two-factor Cairns term structure model using Gibbs sampler with quadratic approximation.

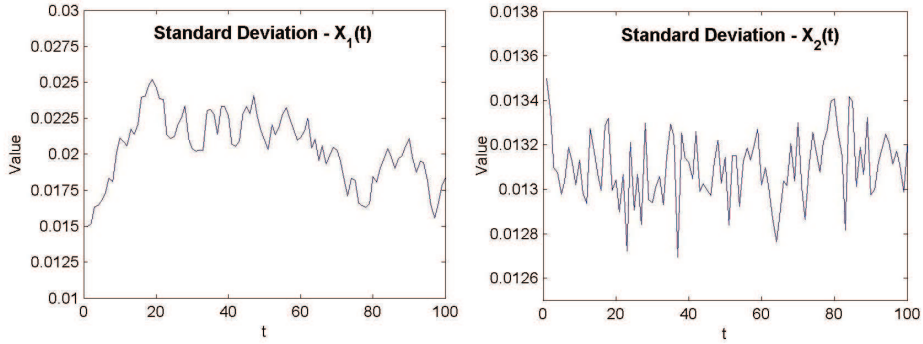

Figure 5.3: Standard deviations of $X_1(t)$(left) and $X_2(t)$(right) for $t = 1, \ldots, 100$ of the two-factor Cairns term structure model using Gibbs sampler with quadratic approximation.

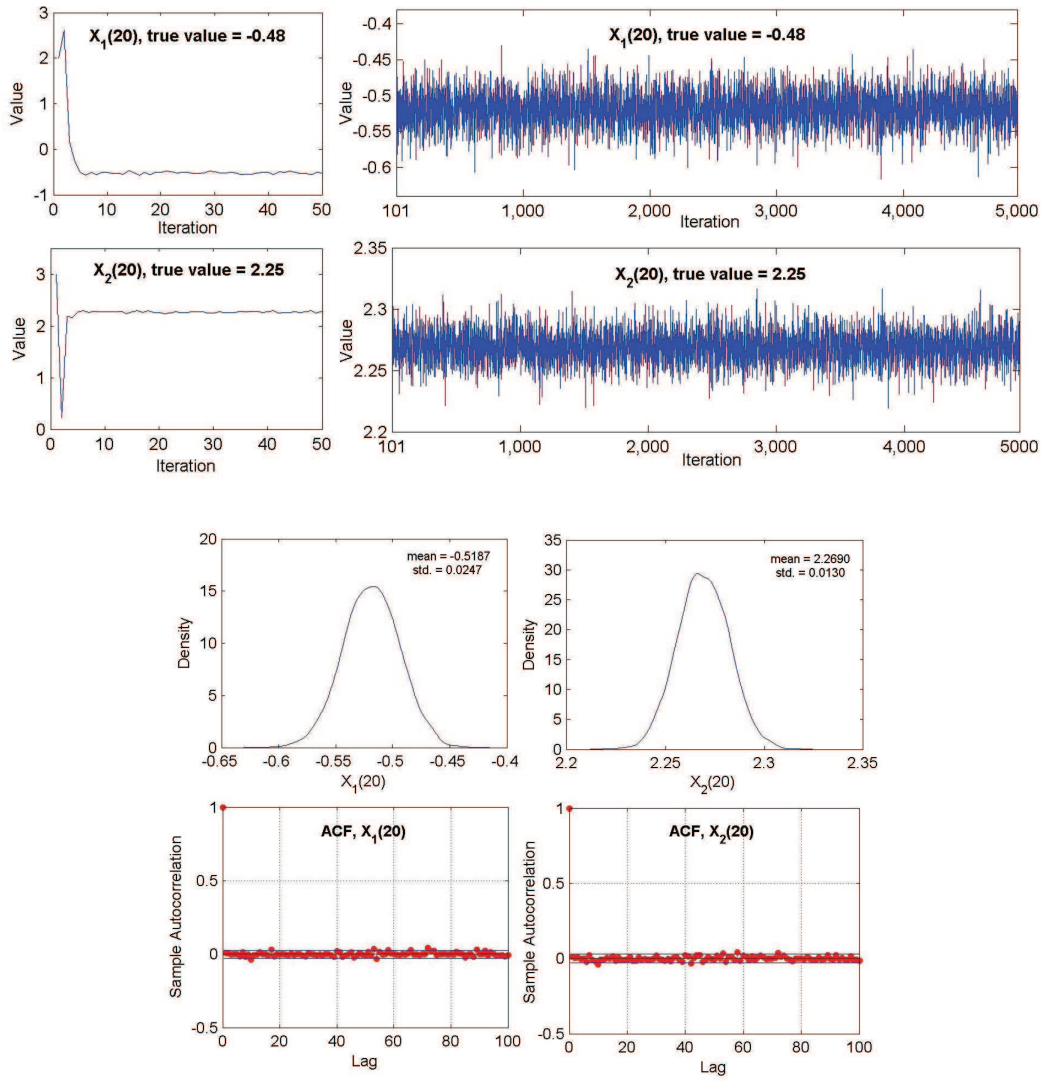

Figure 5.4: Sample paths, posterior densities and autocorrelation functions of $X_1(t)$ and $X_2(t)$ at $t = 20$ of the two-factor Cairns term structure model using Gibbs sampler with quadratic approximation.

5.6.2 Estimating Latent Variables and Model Parameters using the MH Algorithm

In this section, both latent variables and model parameters are estimated using the standard MH algorithm (Algorithm 1) in which the sampling is facilitated by the normal proposal distribution with "constant" variance. Although the Gibbs sampler with quadratic approximation may be employed to search for good initial values for the simulation, we simulate each chain for 15,000 iterations starting from the true values in order to shorten the runtime and particularly concentrate on the convergence assessment.

Figure 5.5 illustrates the sample paths of the model parameters, where γ_1 and γ_2 are fixed to be zero and each parameter is updated individually. From the figure, we can observe that all chains encompass reasonably well the true values but with different velocities of convergence. Compared with the others, σ_1 tends to converge slowest whereas α_1 is found to be most stable and has long excursions in our simulation. Despite the poor convergence of σ_1, the means of all parameters are clearly close to the true values (within one standard deviation) as can be seen from the summary statistics in Table 5.1. Table 5.1 also shows the constant normal proposal standard deviations used for the simulation in which these values are discovered to be suitable for this dataset after extensive tuning up. It should be mentioned that for the model parameters we noticed that the constant proposal standard deviations should be as close to the posterior standard deviations as possible, otherwise the chains will easily drift away since these parameters are too sensitive to be sampled from the proposal with high and low standard deviations. For the MH acceptance rates, we obtain the rates varying from 8.0% to 22.0%. Unsurprisingly, the posterior densities of $\alpha_2, \sigma_1, \sigma_2$ and ρ (Figure 5.6) are not smooth due to weak convergence. Furthermore, the autocorrelation functions of all parameters (Figure 5.7) are slow to decline.

Regarding the latent variables, $X_1(t)$ and $X_2(t)$ are updated as a pair for each t in which we use the constant normal proposal standard deviations 0.055 and 0.04 for all t since it is not practical to tune up the proposal distribution for each t individually. Unlike the model parameters, these values are approximately 1.5 to 2.5 times higher than the resulting posterior standard deviations of all $X_1(t)$ and $X_2(t)$ (the posterior

standard deviations of $X_1(t)$ range from 0.020 to 0.036 and of $X_2(t)$ from 0.013 to 0.019). According to the MH acceptance rates, the pairs of $X_1(t)$ and $X_2(t)$ are accepted with the rates between 5.12% to 10.03% which is rather low.

It is noticed that the latent variables seem to be least sensitive quantities in the Cairns term structure model and hence they are easier controlled than the model parameters. As to the results, the chains strongly converge for almost all t except the unconditional latent variable $X_1(1)$ that is found to be most negatively correlated with σ_1. Figure 5.8 shows the selected sample paths of $X_1(t)$ and $X_2(t)$, for $t = 1, 20, 40, 60, 80$ and 100 (the corresponding plots of posterior densities and autocorrelation functions are also provided in Figures 5.10 and 5.11). Additionally, we can also see from Figure 5.9 that the 95% credible interval constructed from the sample paths of the latent variables include the true values fairly well. This interval is slightly wider than the interval in Figure 5.2 for both $X_1(t)$ and $X_2(t)$, where we estimated the model using Gibbs sampler with quadratic approximation given fixed model parameters.

Next, we conduct further analysis in the interactions among all unknown quantities in this model by considering their cross-correlations (Table 5.2). The scatter plots for the model parameters are shown in Figure 5.12. Among the model parameters, it turns out that there exist strong positive correlations between α_1 and σ_1 and between α_2 and σ_2, while the former is in the lesser degree. Furthermore, β is strongly negatively correlated with a pair of (α_2, σ_2), whereas ρ is least correlated to all other parameters.

Between the model parameters and latent variables, it can be found that α_2 and σ_2 have moderate negative correlation with almost all $X_2(t)$. While α_1 is hardly correlated to $X_1(t)$, σ_1 is found to be most correlated to $X_1(t)$ (can be either positive or negative, with very strong negative correlation with the first $X_1(1), X_1(2), \ldots, X_1(6)$). Furthermore, ρ has consistently positive correlation (around 0.3 to 0.5) to all $X_1(t)$ but almost zero correlation with all $X_2(t)$.

Among all the latent variables $X_1(t)$ and $X_2(t)$ (not shown in the figure), there is no evidence of high correlations except in the group of $X_1(t)$ for $t = 1, 2, \ldots, 6$, in which they are highly positively correlated (around 0.48 to 0.83).

In general, the parameters and latent variables in the same term of the function

$H(u, x)$ in (5.3) tend to be correlated to one another in some way.

The components of the log posterior density are also monitored during the simulation as illustrated in Figure 5.13. As can be seen, although the log-likelihood of the pricing data constitutes of the largest part of the total log posterior, it is most stable and hence we may infer that the variations influencing the overall MH sampling are actually from the log-likelihood of latent variables and priors. The total log posterior of all components has a very similar picture to the log-likelihood of the pricing data but has not been shown here.

In conclusion, we summarise in this second result that tuning for the suitable constant variances of the proposal distribution plays a substantial role in order to obtain good MH convergence. Due to high interaction complexity among the unknown quantities in this model, too small and too large variances of only one parameter can easily make the chain either drift away or diverge. Once one parameter, for example α_2, starts shifting away from the true values to some extent, so do the others (to which α_2 is highly correlated). The total variation of the simulation is hard to be controlled by using the constant proposal variance and much relies on the variation of the log-likelihood of latent variables.

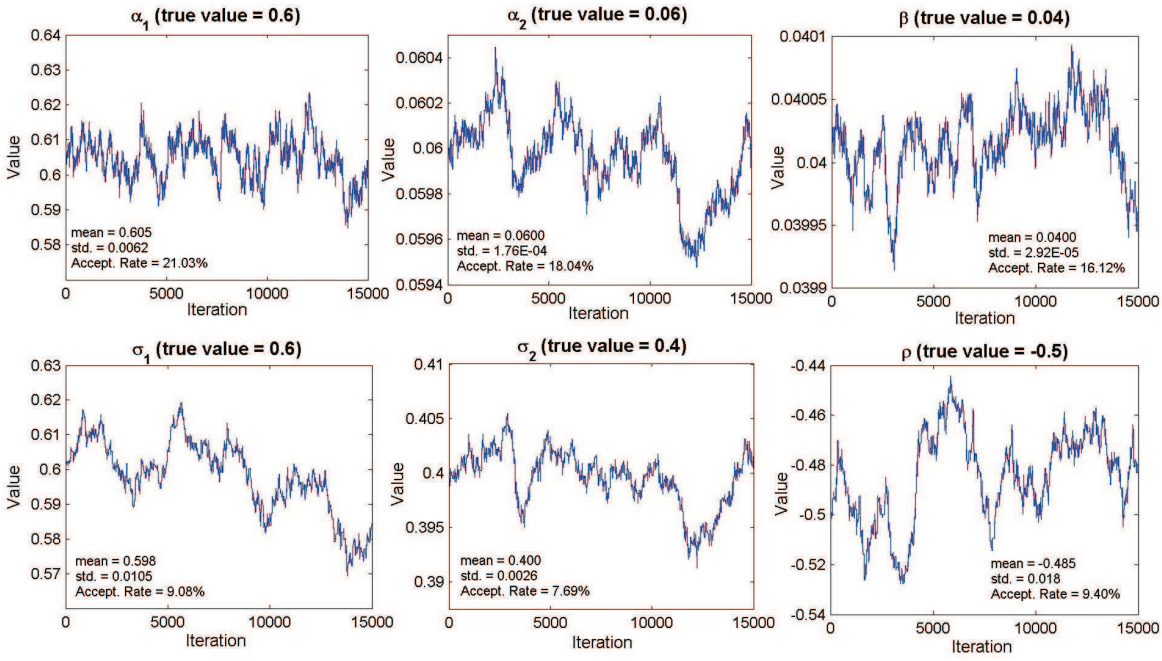

Figure 5.5: Sample paths of model parameters of the two-factor Cairns term structure model using the MH algorithm with constant normal proposal variance.

	(True value)	Mean	Std.	95% Credible Interval	Proposal std.	Acceptance rate
β	(0.04)	0.0400	0.00003	(0.03995, 0.04007)	0.00003	16.12%
α_1	(0.6)	0.605	0.0062	(0.5921, 0.6158)	0.0060	21.03%
α_2	(0.06)	0.0600	0.00018	(0.05956, 0.06026)	0.00010	18.04%
σ_1	(0.6)	0.598	0.0104	(0.5756, 0.6150)	0.0100	9.08%
σ_2	(0.4)	0.400	0.0026	(0.3935, 0.4038)	0.0050	7.69%
ρ	(-0.5)	-0.485	0.0180	(-0.5218, -0.4555)	0.0200	9.40%

Table 5.1: Summary statistics of parameter posterior estimates of the two-factor Cairns term structure model using the MH algorithm with constant normal proposal variance (15,000 iterations).

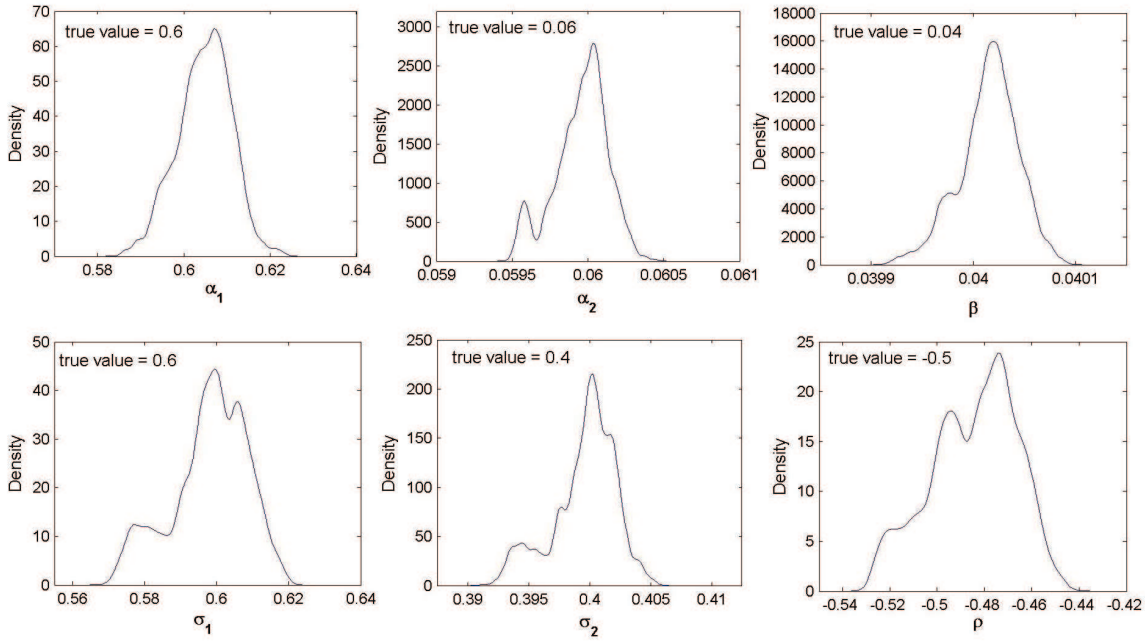

Figure 5.6: Posterior densities of model parameters of the two-factor Cairns term structure model using the MH algorithm with constant normal proposal variance.

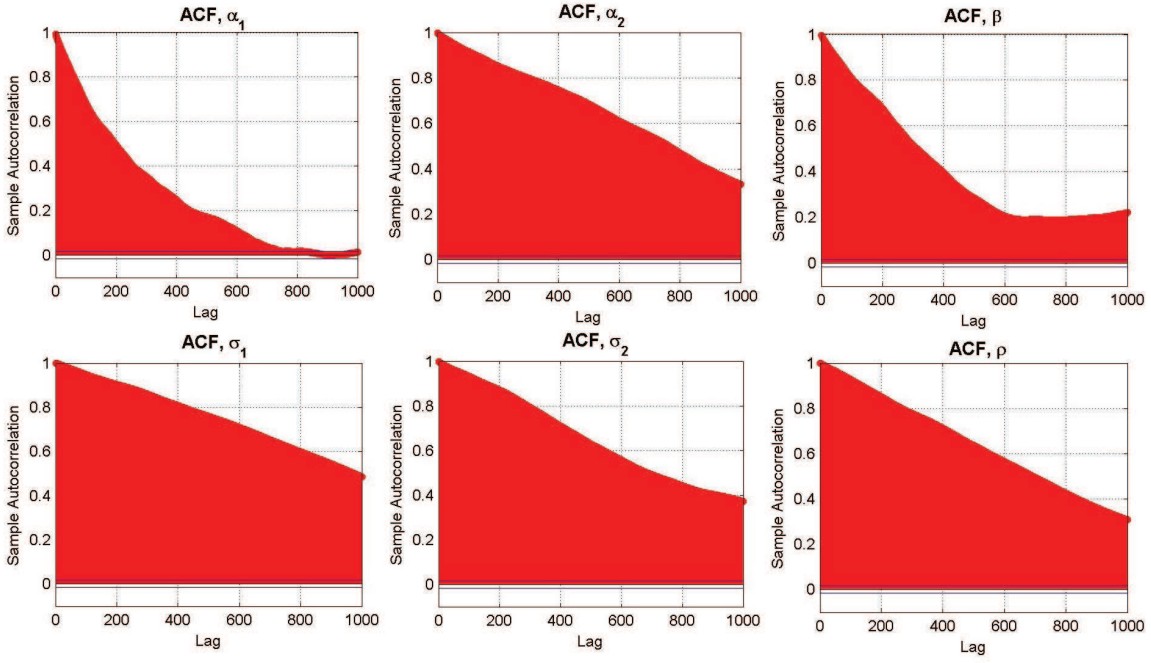

Figure 5.7: Autocorrelation functions of model parameters of the two-factor Cairns term structure model using the MH algorithm with constant normal proposal variance.

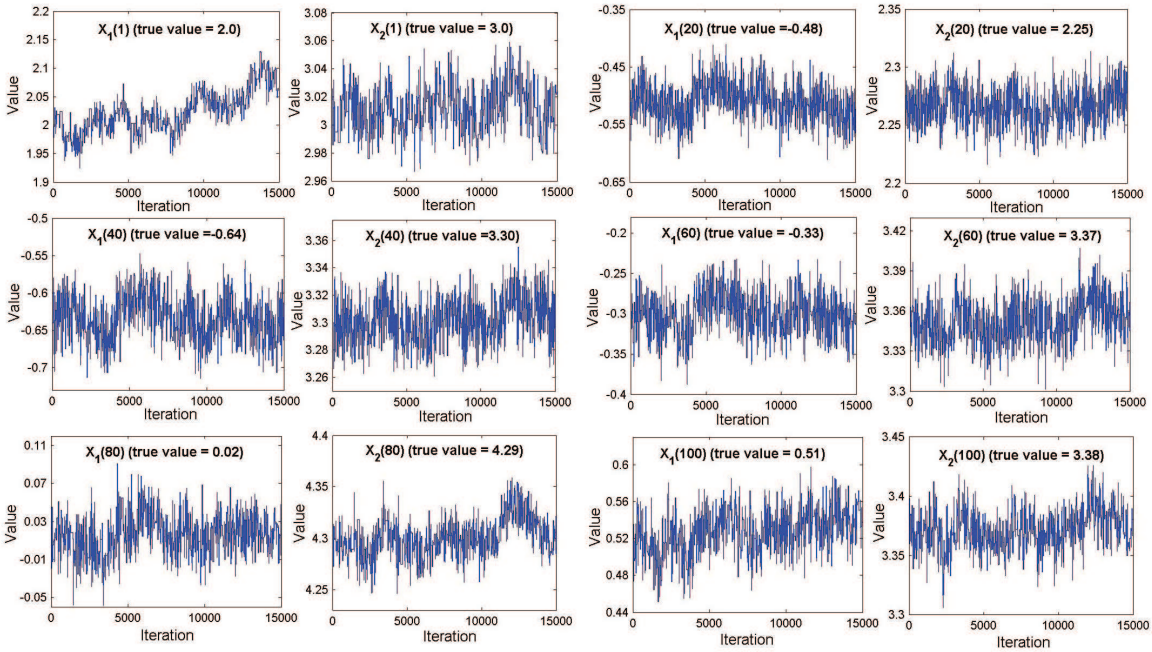

Figure 5.8: Sample paths of latent variables (for $t = 1, 20, 40, 60, 80$ and 100) of the two-factor Cairns term structure model using the MH algorithm with constant normal proposal variance.

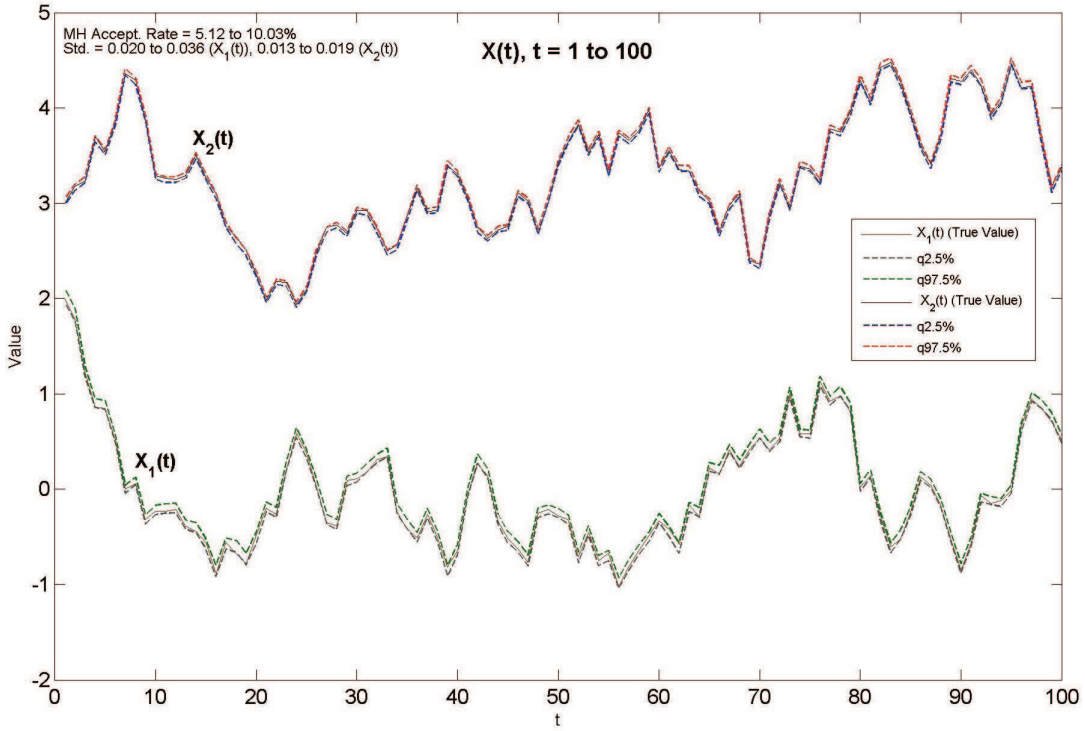

Figure 5.9: Plots of 95% credible interval constructed from the sample paths with the true values of $X_1(t)$ and $X_2(t)$ for $t = 1, \ldots, 100$, of the two-factor Cairns term structure model using the MH algorithm with constant normal proposal variance.

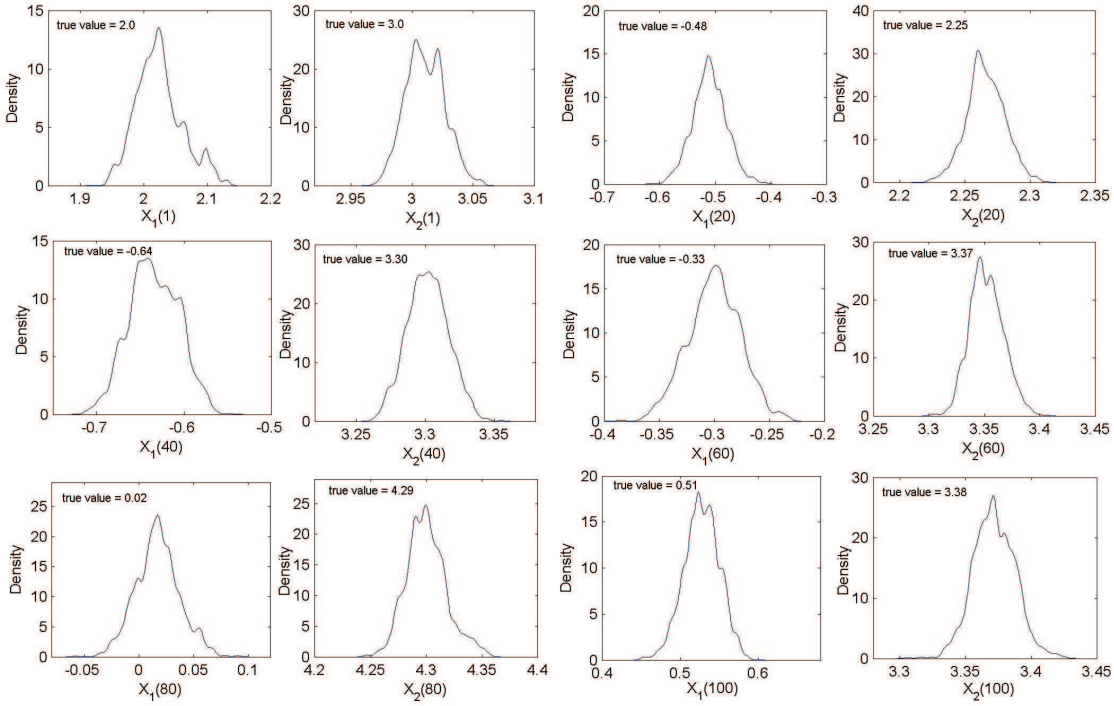

Figure 5.10: Posterior densities of latent variables (for $t = 1, 20, 40, 60, 80$ and 100) of the two-factor Cairns term structure model using the MH algorithm with constant normal proposal variance.

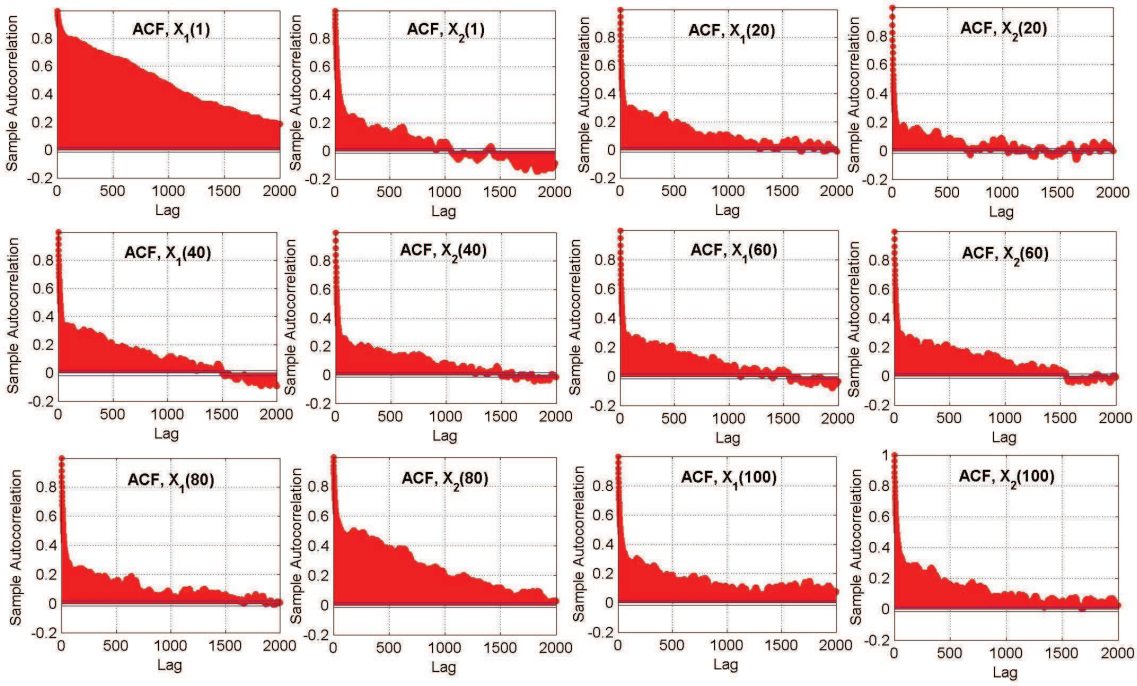

Figure 5.11: Autocorrelation functions of latent variables (for $t = 1, 20, 40, 60, 80$ and 100) of the two-factor Cairns term structure model using the MH algorithm with constant normal proposal variance.

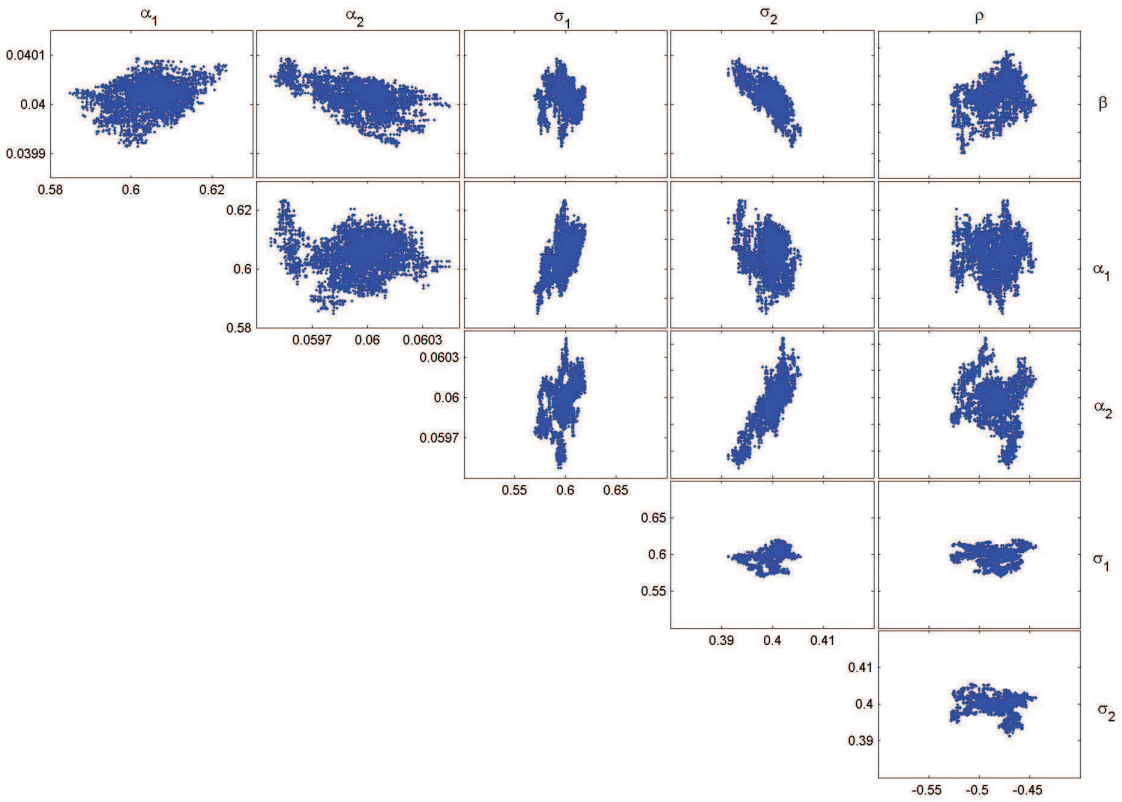

Figure 5.12: Scatter plots of model parameters of the two-factor Cairns term structure model using the MH algorithm with constant normal proposal variance.

	β	α_1	α_2	σ_1	σ_2	ρ	$X_1(t)$	$X_2(t)$
β	1.00	0.28	-0.54	-0.08	-0.71	0.37	0.00 to 0.30	-0.37 to 0.35
α_1		1.00	0.01	0.57	-0.19	0.02	-0.42 to 0.22	-0.42 to 0.32
α_2			1.00	0.39	0.85	-0.16	-0.38 to 0.21	-0.68 to 0.02
σ_1				1.00	0.31	-0.02	-0.86 to 0.52	-0.86 to 0.02
σ_2					1.00	-0.18	-0.30 to 0.23	-0.68 to 0.17
ρ						1.00	0.29 to 0.55	-0.22 to 0.31

Table 5.2: Correlation matrix of model parameters and latent variables of the two-factor Cairns term structure model using the MH algorithm with constant normal proposal variance.

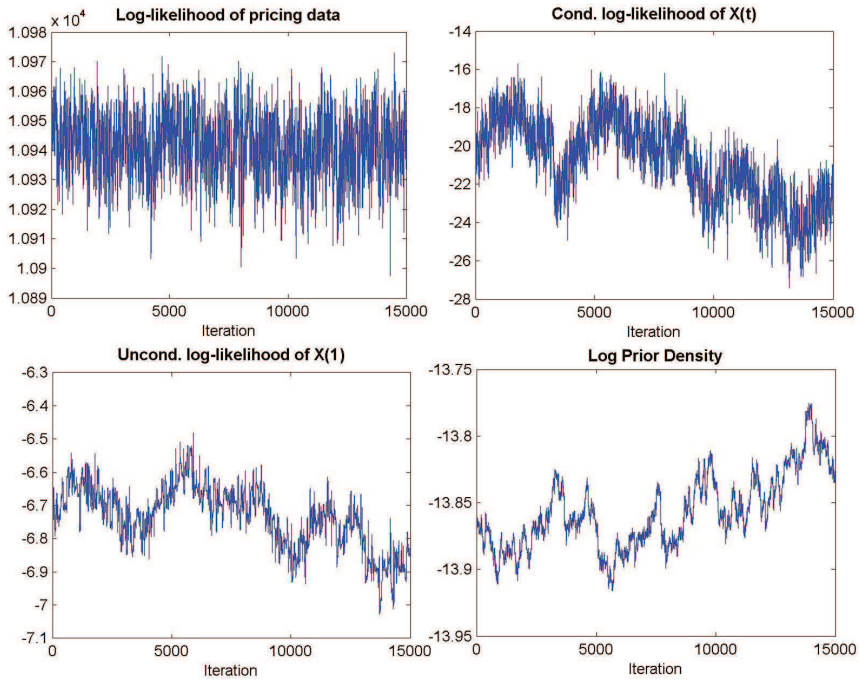

Figure 5.13: Log posterior components (referring to equation 5.9): log-likelihood of pricing data (top, left), conditional and unconditional log-likelihood of latent variables (top, right and bottom, left) and log prior density (bottom, right) using the MH algorithm with constant normal proposal variance.

5.6.3 Estimating Latent Variables and Model Parameters using the MH Algorithm with Some Improvements

Although the chains according to the previous result tend to converge slowly, they result in reasonable inference for all the model parameters and latent variables. Nonetheless, some improvements are yet required in order to implement the MH algorithm with more complex data. Having the standard MH algorithm with constant normal proposal variance (Algorithm 1) as a base case, we here consider to improve the proposal distribution using the adaptive Metropolis-Hastings algorithm (Algorithm 3) associated with a blocking strategy, reparameterising the log posterior of Cairns bond prices and re-evaluating the prior distributions. All of these will be first described below, before we move to look at the results.

5.6.3.1 Reparameterising

From Table 5.2, there is evidence of strong correlation between some parameters, particularly σ_1 and σ_2 where they are highly correlated to $X_1(t)$ and $X_2(t)$ for all t respectively. In the bond price formula, σ_1 and σ_2 are actually the local volatilities of all the latent variables $X_1(t)$ and $X_2(t)$. Thus, here we attempt to eliminate such correlations by re-parameterisation.

Let $Y(t) = (Y_1(t), Y_2(t))'$, where $Y_1(t) = \sigma_1 X_1(t)$ and $Y_2(t) = \sigma_2 X_2(t)$ and $\gamma_y = (\gamma_{y_1}, \gamma_{y_2})'$, where $\gamma_{y_1} = \sigma_1\gamma_1$ and $\gamma_{y_2} = \sigma_2\gamma_{y_2}$. Then, the log posterior in (5.10) can be re-written as

$$\begin{aligned}\log f(\Theta|\mathbb{P}) \quad \propto \quad & -\frac{MN_t}{2}\log(2\pi\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2}\sum_{t=1}^{M}\sum_{j=1}^{N_t}\left(P(t,\tau_{tj}) - C_Y(\tau_{tj}; Y(t), \theta)\right)^2 \\ & -M\log(2\pi) - \frac{1}{2}\log|\Omega_y| - \frac{(M-1)}{2}\log|\Sigma_Y| \\ & -\frac{1}{2}(Y(1)-\gamma_y)'\Omega_y^{-1}(Y(1)-\gamma_y) - \frac{1}{2}\sum_{t=2}^{M}\hat{Z}_Y(t)'\Sigma^{-1}\hat{Z}_Y(t) + \log f_0(\theta), \end{aligned} \tag{5.21}$$

where

$$C_Y(\tau, y, \theta) = \frac{\int_\tau^\infty H(u,y)du}{\int_0^\infty H(u,y)du},$$

$$H(u,y) = \exp\Big[-\beta u + y_1\, e^{-\alpha_1 u} + y_2\, e^{-\alpha_2 u} - \frac{1}{2}\sum_{i,j=1}^{2}\frac{\rho_{ij}\sigma_i\sigma_j}{\alpha_i+\alpha_j}e^{-(\alpha_i+\alpha_j)u}\Big]$$

$$\hat{Z}_Y(t) = Y(t) - \gamma_y - K(Y(t-1) - \gamma_y),$$

$$\Sigma_Y = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \frac{\sigma_1^2}{2\alpha_1}(1-e^{-2\alpha_1\Delta t}) & \frac{\rho\sigma_1\sigma_2}{\alpha_1+\alpha_2}(1-e^{-(\alpha_1+\alpha_2)\Delta t}) \\ \frac{\rho\sigma_1\sigma_2}{\alpha_1+\alpha_2}(1-e^{-(\alpha_1+\alpha_2)\Delta t}) & \frac{\sigma_2^2}{2\alpha_2}(1-e^{-2\alpha_2\Delta t}) \end{pmatrix},$$

$$\Omega_y = \begin{pmatrix} \frac{\sigma_{11}}{1-k_1^2} & \frac{\sigma_{12}}{1-k_1k_2} \\ \frac{\sigma_{21}}{1-k_1k_2} & \frac{\sigma_{22}}{1-k_2^2} \end{pmatrix}, K = \begin{pmatrix} e^{-\alpha_1\Delta t} & 0 \\ 0 & e^{-\alpha_2\Delta t} \end{pmatrix}.$$

5.6.3.2 Adaptive MH Algorithm and Blocking Strategy

Previously, we made an observation that the normal proposal distribution with constant variance may not facilitate well the MH algorithm for the model parameters. Moreover, it does spend long time to explore suitable variance values. As already described in an earlier section, one way to improve this is to use the adaptive Metropolis-Hastings algorithm (Algorithm 3) which is also associated with a blocking strategy. First of all, based on some evidence of the correlation structure, we group the model parameters and latent variables as follows.

I. α_1, σ_1 and ρ.

II. α_2, σ_2 and β.

III. γ_1 and γ_2.

IV. $X_1(t)$ and $X_2(t)$ for each t.

In each group, the parameters or latent variables will be updated together in the MH algorithm. For our simulation, candidate points for the parameters in group I and II will be sampled from a multivariate normal distribution where the means depend on their previous values with the covariance matrix computed from the most recent previous 200 values of sample paths of the parameters in the group (an arbitrary fixed covariance matrix will be used for the first 200 iterations). For example, at 601st iteration, the covariance matrix of the proposal distribution is of sample paths from 401st to 600th iteration, at 602nd iteration from 402nd to 601st iteration and so on. Similarly, candidate points for the parameters and latent variables in group III and IV, will be generated by the same way as the first two groups except we will impose zero correlation to the proposal covariance matrix at all time.

5.6.3.3 Re-evaluating the Priors

Finally, we allow the priors of the parameters $\beta, \alpha_1, \alpha_2, \sigma_1$ and σ_2 to be slightly more informative. We set their means with reference to the posterior means of the earlier simulations and reduce the coefficient of variation to around 5.3 (previously a $\Gamma(0.01, 100)$ prior, which provides a mean of 1.0 and coefficient of variation of 10, was assigned to all of these parameters). Specifically, the priors of all the model

parameters now become

$$\begin{aligned} f_0(\beta) &\sim \Gamma(0.036, 1.13), \\ f_0(\alpha_1), f_0(\sigma_1) &\sim \Gamma(0.036, 16.67), \\ f_0(\alpha_2) &\sim \Gamma(0.036, 1.67), \\ f_0(\sigma_2) &\sim \Gamma(0.036, 11.25), \\ f_0(\rho) &\sim U[-1, 1], \\ f_0(\gamma_1), f_0(\gamma_2) &\sim N(0, 1.0 \times 10^5). \end{aligned} \tag{5.22}$$

5.6.3.4 Results with the Improvements

To begin with, we provide a narrative summary of the effects for each improvement from several simulations compared with the base case as follows.

- We first started using adaptive proposal distributions for each parameter individually, but it was not possible to notice any distinct difference according to the results.

- Then, the reparameterisation for the latent variables $X_1(t)$ and $X_2(t)$ (i.e. define $Y_1(t) = \sigma_1 X_1(t), Y_2 = \sigma_2 X_2(t)$) was therefore considered and it turned out that the convergence of σ_1 was clearly improved but σ_2 still converged slower than expected.

- Next, we re-evaluated the priors for $\beta, \alpha_1, \alpha_2, \sigma_1, \sigma_2$ as described in the earlier section. That is, their prior means were shifted from 1.0 to the values with respect to their posterior means with variances of 10.0. However, any difference was hard to notice. Consequently, we also decreased the coefficient variations of the priors to around 5.3 but once again distinct improvement could not be observed.

- Eventually, we attempted to improve the convergence by incorporating the blocking strategy. In effect, we found that the convergence of σ_2 was evidently better.

In the following results, we run MCMC simulation for 20,000 iterations for each chain using the adaptive MH algorithm with the reparameterised log posterior in (5.21) and the priors in (5.22) where γ_{y_1} and γ_{y_2} are now unrestricted and then

estimated. According to the resulting sample paths shown in Figure 5.14, we can easily see that the convergence of the parameters σ_1, α_2 and β are all clearly improved (compared with Figure 5.5). Parameter α_1 still converges well same as the previous result, while ρ is not significantly different. Furthermore, it can also be noticed that the posterior standard deviations (Table 5.3) are relatively higher (about twice than before). In addition, γ_{y_1} and γ_{y_2} undoubtedly get stationary although the range of γ_{y_2} is rather wide.

Plots of posterior densities and autocorrelation functions are illustrated in Figures 5.15 and 5.16. Comparing to Figures 5.6 and 5.7 for the standard MH algorithm, the shapes of the densities of almost all parameters (except α_1) are clearly smoother and the autocorrelation functions are generally improved (again, those of α_1 are fairly similar to those previously presented).

Figures 5.17 and 5.18 present the adaptive proposal standard deviations and correlations of the parameters in group I $(\alpha_1, \sigma_1, \rho)$ and II $(\alpha_2, \sigma_2, \beta)$ respectively. In the simulation, we use the constant covariances for the initial 200 iterations and afterwards covariances are computed based on their previous most recent 200 values and scaled up by appropriate factors as shown in the last column in Table 5.3. From the figures, we can notice that the proposal standard deviations are no longer constant but fluctuating around some certain values in all cases. For the proposal correlations, we found that while the values for the parameters in group I spread throughout the range of $(-1, 1)$, those for the parameters in group II are relatively more stable and highly positive (tend to get close to 1). Perhaps, this strongly positive correlation structure is caused by the reparametersation of the log posterior.

For the latent variables, the sample paths of $Y_1(t)$ and $Y_2(t)$, for $t = 1, 20, 40, 60, 80,$ and 100, are demonstrated in Figure 5.19. As can be noticed, the convergence of $Y_1(1)$ compared with $X_1(1)$ in Figure 5.8 is evidently better, whereas the remaining still converge well. Figure 5.20 demonstrates plots of the 95% credible interval constructed from the sample paths of $Y_1(t)$ and $Y_2(t)$ for all t. Comparing to those of $X_1(t)$ and $X_2(t)$ in Figure 5.9, these intervals are generally wider which infers higher posterior standard deviations. Regarding the posterior densities and autocorrelation functions (Figures 5.21 and 5.22), significant improvement can be observed for $Y_1(1)$ which is reparameterised from $X_1(1)$.

Figures 5.23 and 5.24 show the adaptive proposal standard deviations of selected $Y_1(t)$ and $Y_2(t)$, where these are scaled up by 2.5 and 2.0 respectively. From the figures, we can see that the movement behaviours are similar to those for the model parameters (Figures 5.17 and 5.18). The advantage of using adaptive proposal variance for the latent variables is that currently the proposals are adaptive for each t, whereas in the past we used the same constant proposal variance values for all t. Regarding the acceptance MH rates, we achieve rates lying between 11.51% to 14.25%.

Next, we investigate the correlation structure of the model parameters and latent variables. From the scatter plots in Figure 5.25, we can observe that overall the correlation structure of the model parameters is much better than those obtained with the standard MH algorithm (Figure 5.12). However, we observe strongly positive correlations among all parameters in group II ($\alpha_2, \sigma_2, \beta$). In the previous result (Table 5.2), although β was found to be strongly negatively correlated to α_2 and σ_2, we found that this was not always the case. Specifically, since the chains converged rather slowly, their estimated correlations are much less reliable than those after the reparameterisation in which we can easily see that the chains generally converge much faster.

Between the model parameters and latent variables, it turns out that ρ is most likely to correlate with $Y_1(t)$ for all t. Among the latent variables themselves, there is no evidence of any strong correlation among them. Moreover, the strong negative correlations previously found for the first $X_1(1), X_1(2), \ldots, X_1(6)$ now disappear.

Additionally, we also compare the log posterior components monitored during the simulation as shown in Figure 5.26 to Figure 5.13. While the log-likelihood of the pricing data roughly remains unchanged, the variations of other components are much more stable. This somehow corresponds to the better convergence achieved for both model parameters and latent variables.

In the end, we conclude for this chapter that the achieved results using the adaptive MH algorithm with a blocking strategy and reparameterising the log posterior distribution were substantially improved from those with the standard MH algorithm, in terms of both convergence and correlation structure. Nevertheless, there was no significant improvement resulting from the use of the re-evaluated priors. Further-

more, we observed that the chains of almost all parameters and latent variables can achieve stationarity much easier than using the standard MH algorithm. Although parameter ρ is still rather sensitive to the proposal variance and hard to converge, this is not a surprising result since ρ appears in the minor term in the bond price posterior which tends to be most difficult to be estimated accurately. At this point, the algorithm seems to be efficient enough that we can proceed to fit the model to real market data. We do this in the next chapter.

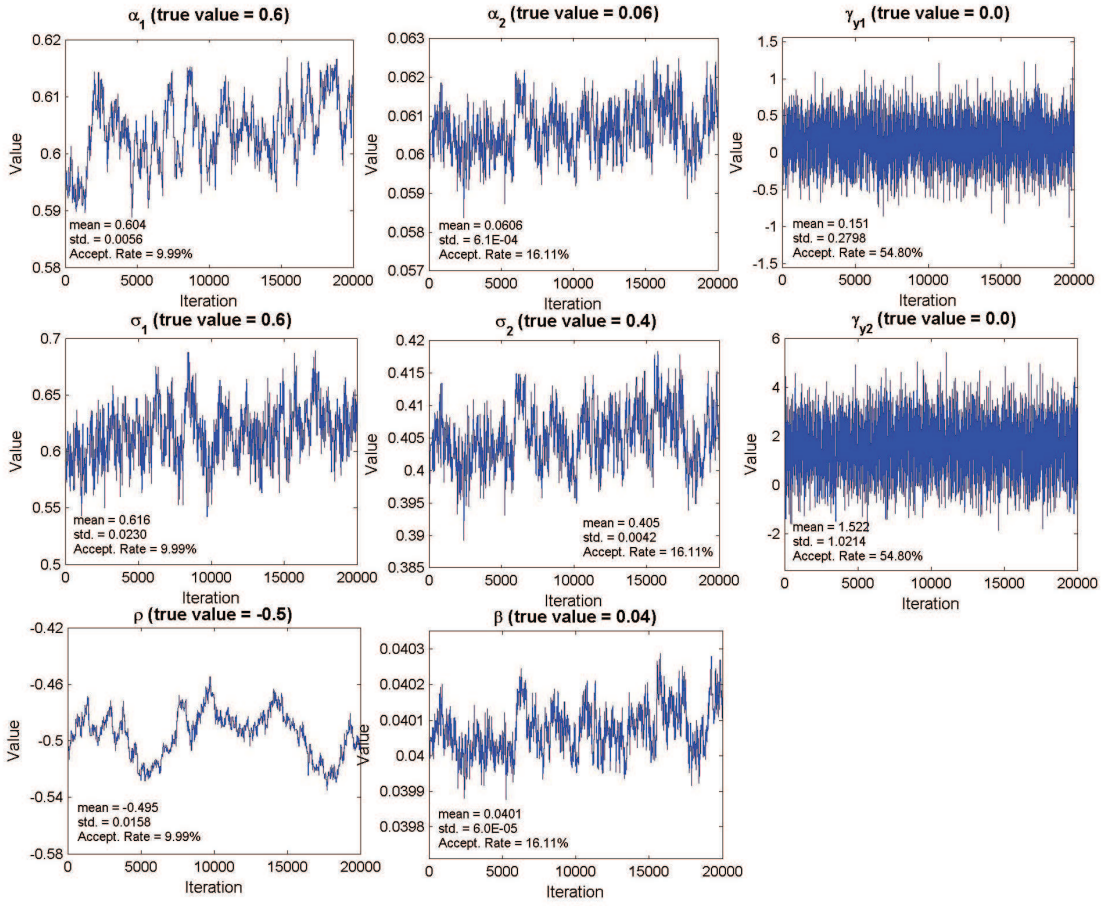

Figure 5.14: Sample paths of model parameters of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors.

	(True value)	Mean	Std.	95% Credible Interval	Acceptance rate	Scaling of the proposal std.
α_1	(0.6)	0.604	0.0056	(0.5924, 0.6143)	9.99%	2.0
σ_1	(0.6)	0.616	0.0230	(0.5688, 0.6616)	9.99%	2.0
ρ	(-0.5)	-0.495	0.0158	(-0.5250, -0.4673)	9.99%	2.0
α_2	(0.06)	0.0606	0.00061	(0.05958, 0.06190)	16.11%	1.4
σ_2	(0.4)	0.405	0.0042	(0.3977, 0.4138)	16.11%	1.4
β	(0.04)	0.0401	0.00006	(0.03996, 0.04021)	16.11%	1.4
γ_{y_1}	(0.0)	0.151	0.2798	(-0.3976, 0.7117)	54.80%	1.0
γ_{y_2}	(0.0)	1.522	1.0214	(-0.5068, 3.4789)	54.80%	1.0

Table 5.3: Summary statistics of parameter posterior estimates of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors (20,000 iterations).

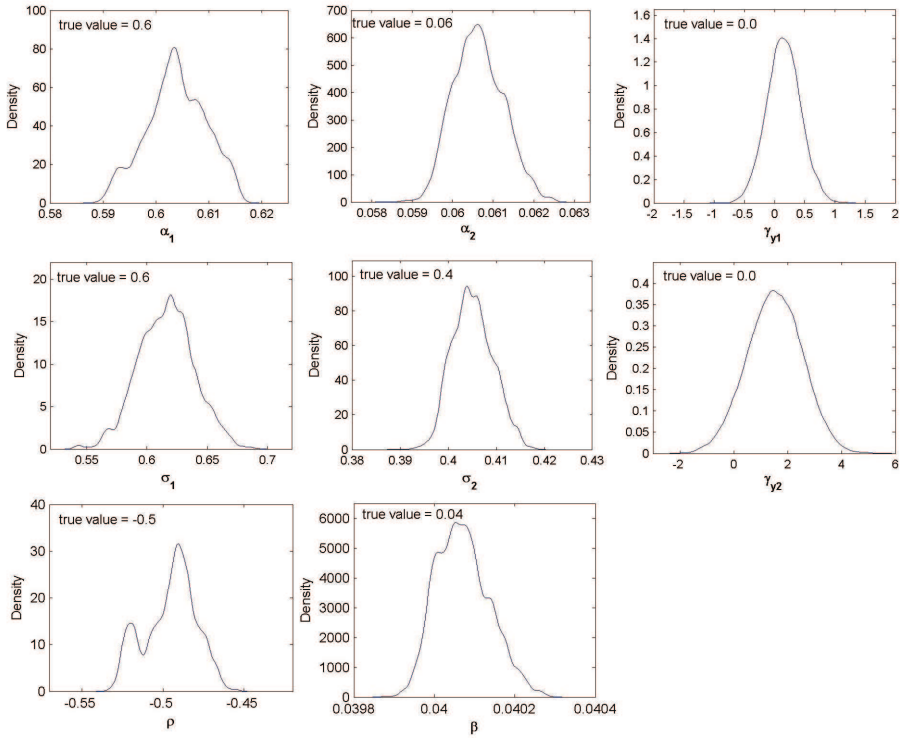

Figure 5.15: Posterior densities of model parameters of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors.

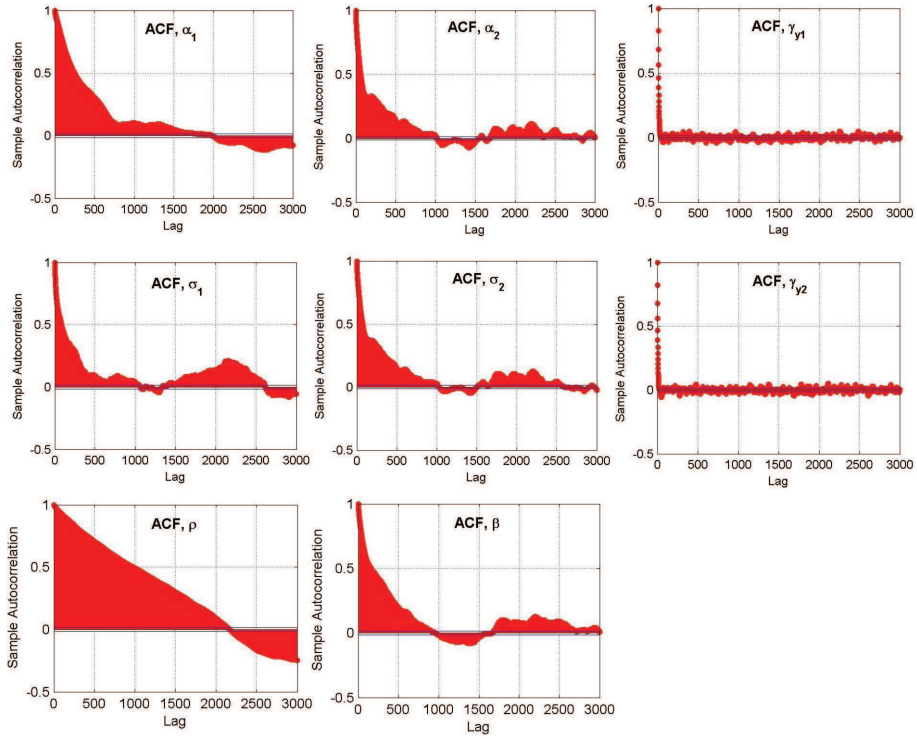

Figure 5.16: Autocorrelation functions of model parameters of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors.

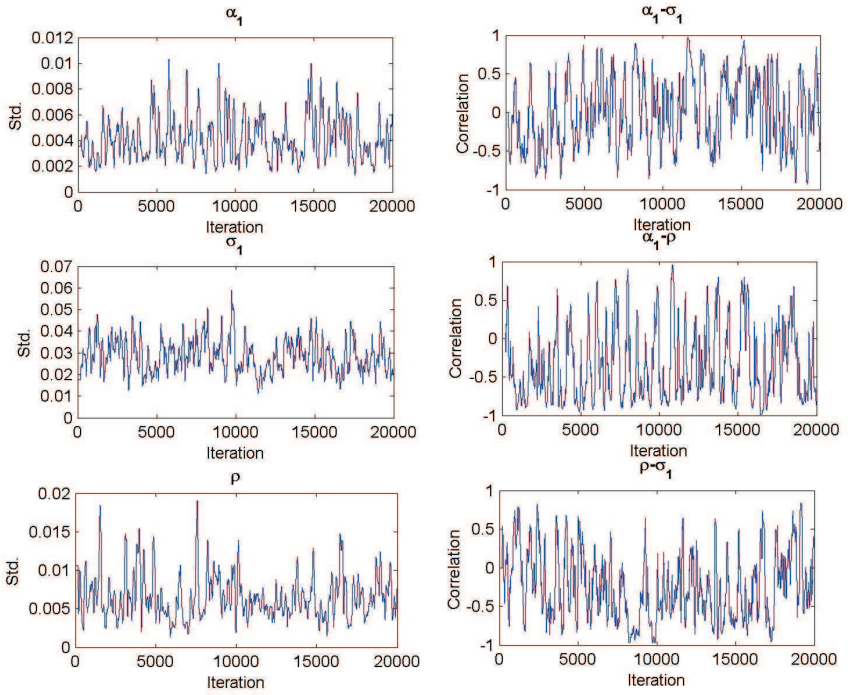

Figure 5.17: Proposal standard deviations (left column) and correlation (right column) by the adaptive MH algorithm for the parameters α_1, σ_1 and ρ. The values are computed from their most recent previous 200 values.

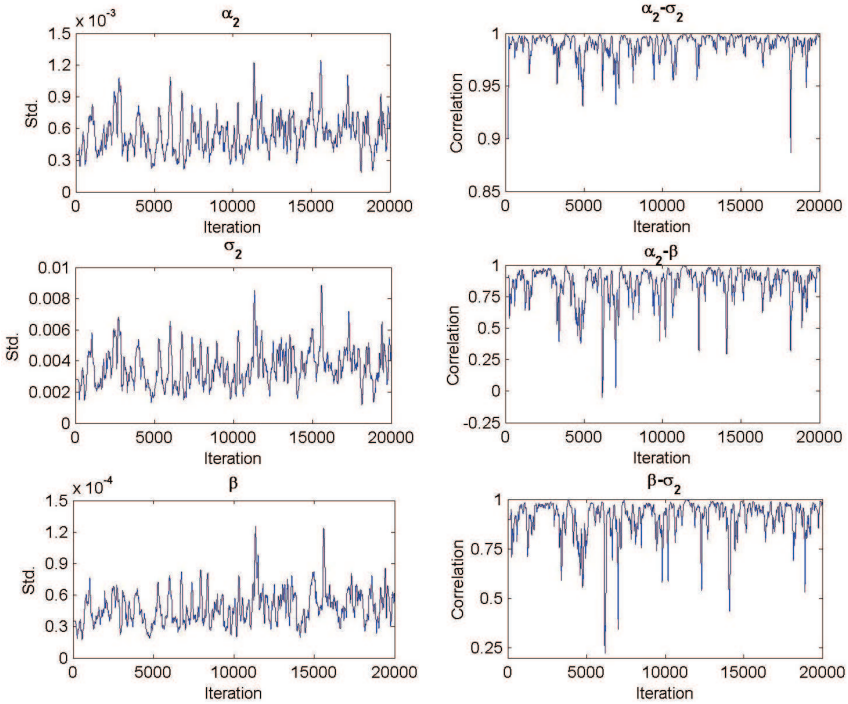

Figure 5.18: Proposal standard deviations (left column) and correlation (right column) by the adaptive MH algorithm for the parameters α_2, σ_2 and β. The values are computed from their most recent previous 200 values.

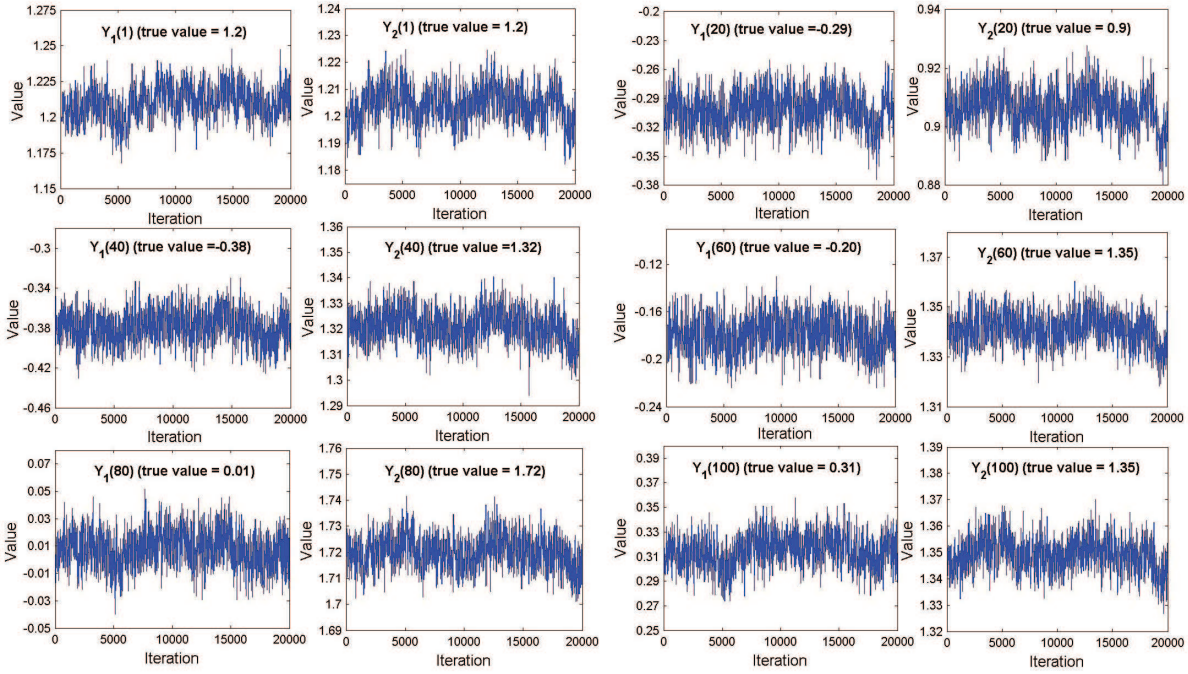

Figure 5.19: Sample paths of latent variables (for $t = 1, 20, 40, 60, 80$ and 100) of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors.

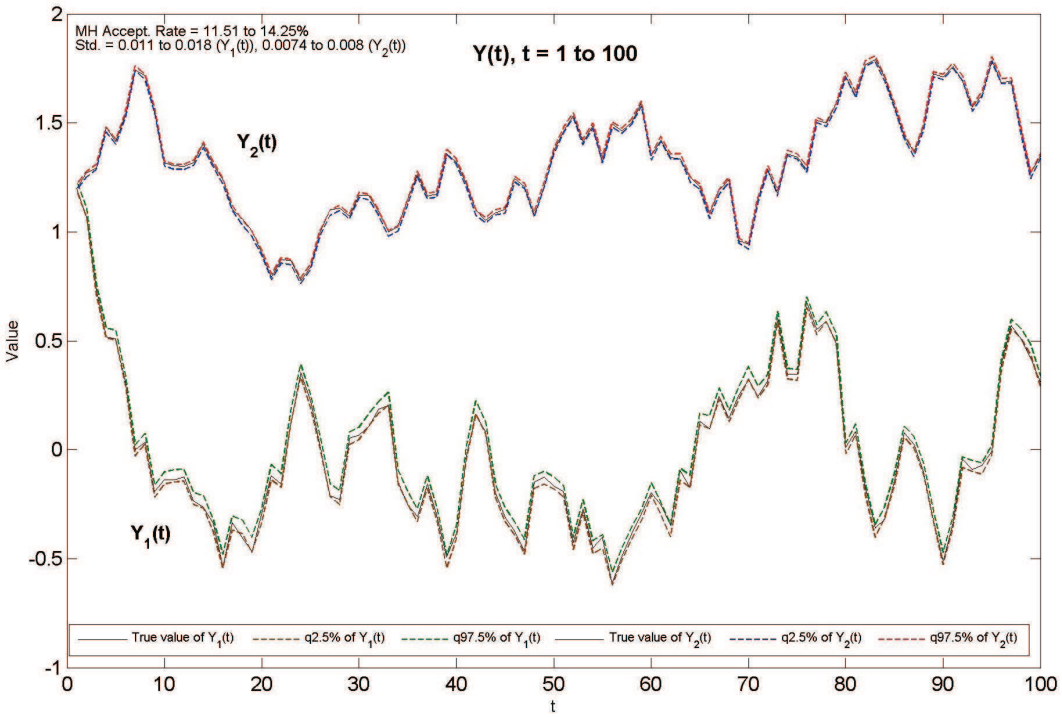

Figure 5.20: Plots of 95% credible interval constructed from the sample paths with the true values of $Y_1(t)$ and $Y_2(t)$ for $t = 1, \ldots, 100$, of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors.

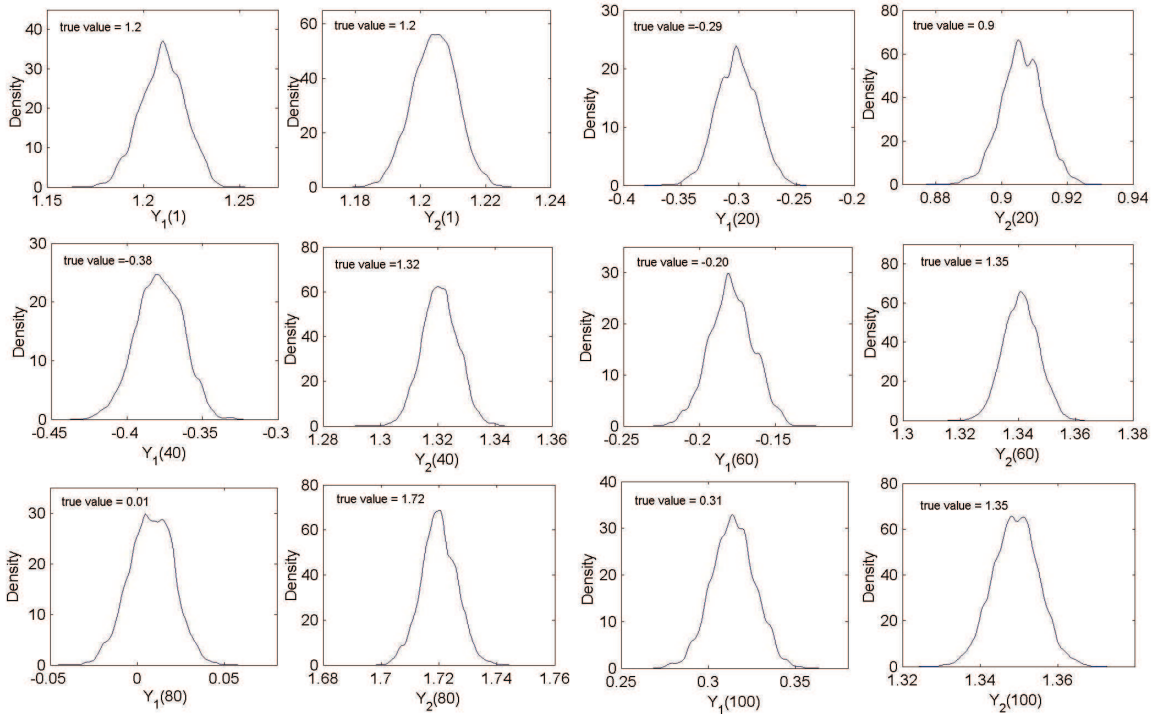

Figure 5.21: Posterior densities of latent variables (for $t = 1, 20, 40, 60, 80$ and 100) of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors.

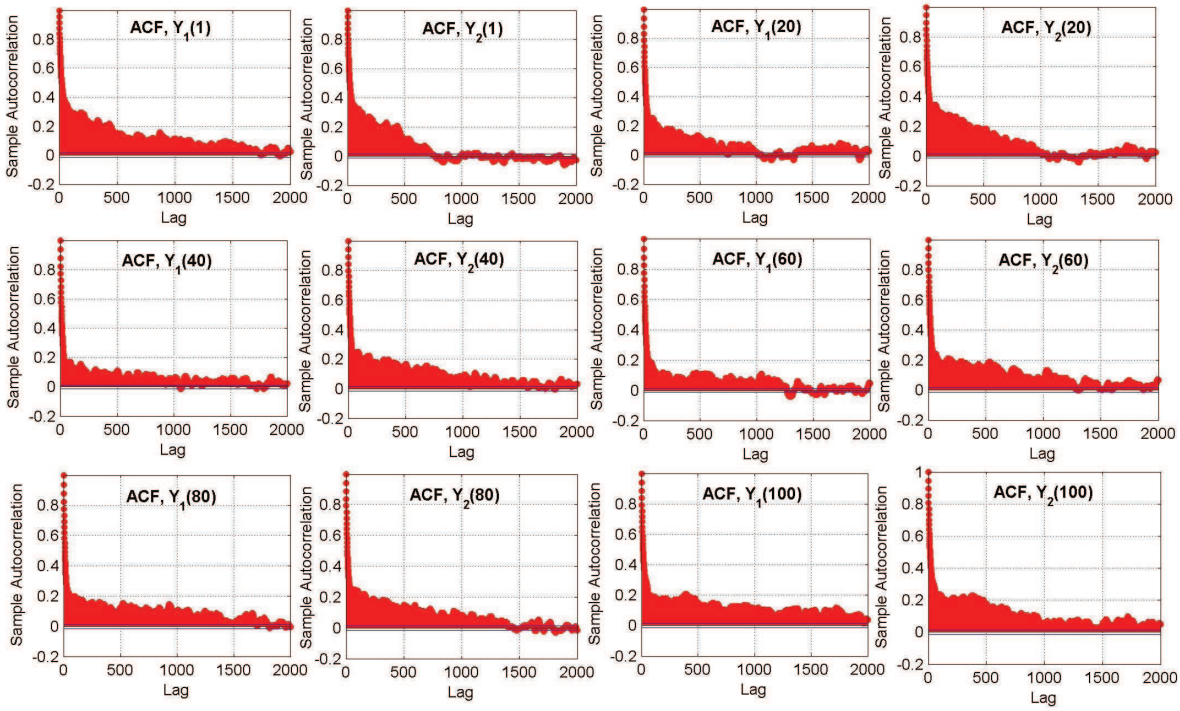

Figure 5.22: Autocorrelation functions of latent variables (for $t = 1, 20, 40, 60, 80$ and 100) of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors.

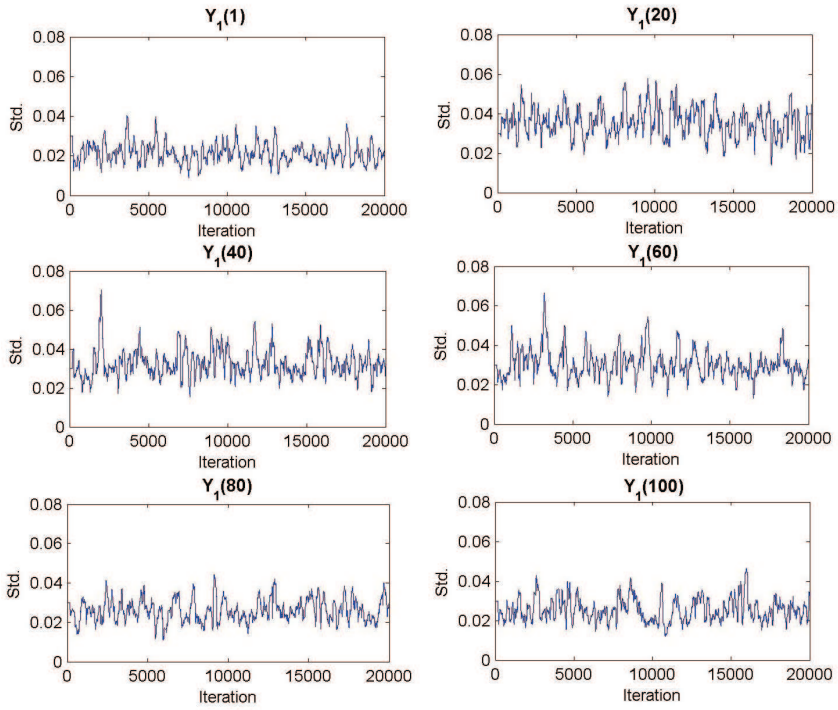

Figure 5.23: Proposal standard deviations by the adaptive MH algorithm for the latent variables $Y_1(t)$ (for $t = 1, 20, 40, 60, 80$ and 100). The values are computed from their most recent previous 200 values.

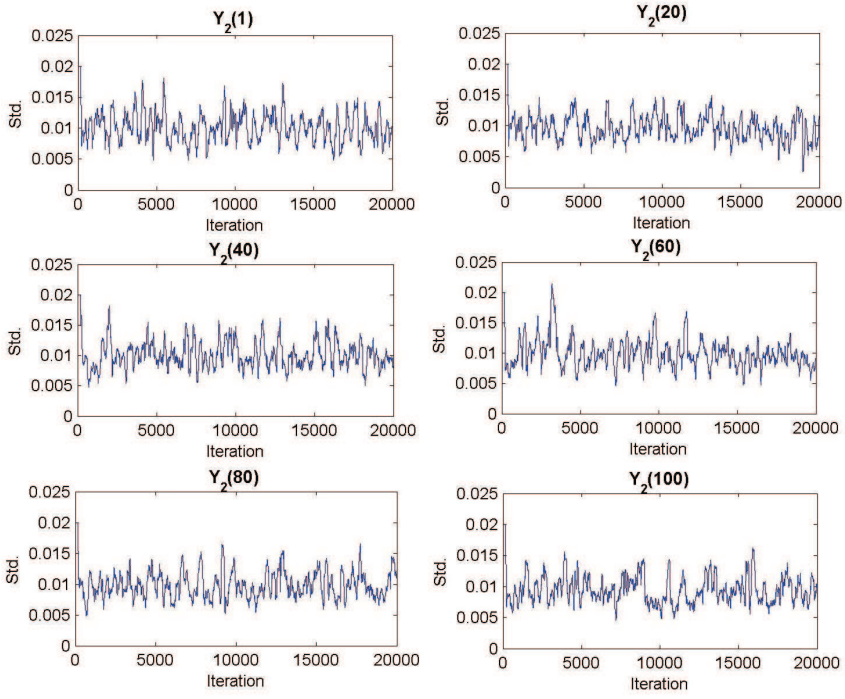

Figure 5.24: Proposal standard deviations by the adaptive MH algorithm for the latent variables $Y_2(t)$ (for $t = 1, 20, 40, 60, 80$ and 100). The values are computed from their most recent previous 200 values.

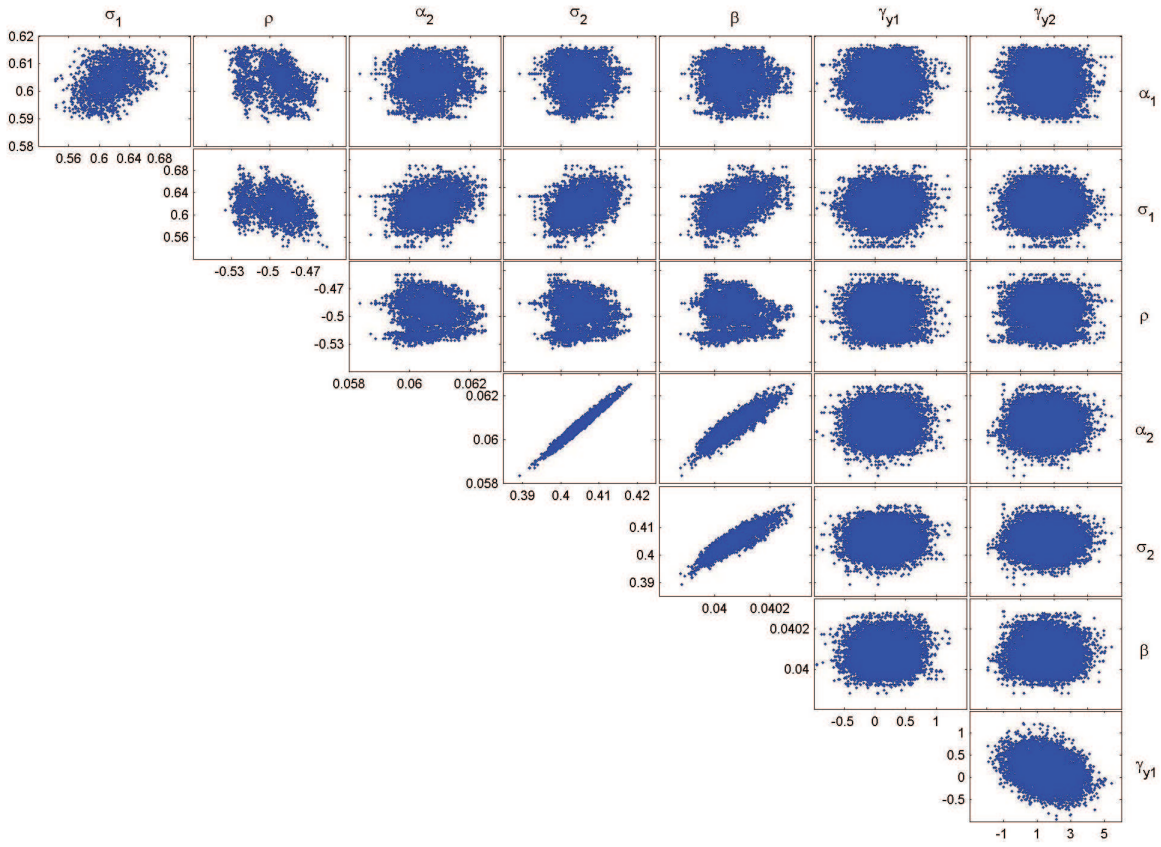

Figure 5.25: Scatter plots of model parameters of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors.

	α_1	σ_1	ρ	α_2	σ_2	β	γ_{y_1}	γ_{y_2}	$Y_1(t)$	$Y_2(t)$
α_1	1.00	0.37	-0.27	0.00	-0.02	0.03	0.01	-0.03	-0.23 to 0.20	-0.20 to 0.29
σ_1		1.00	-0.34	0.29	0.35	0.43	0.05	-0.01	-0.10 to 0.29	-0.22 to 0.29
ρ			1.00	-0.02	-0.05	-0.03	-0.01	0.02	0.24 to 0.44	-0.10 to 0.32
α_2				1.00	0.98	0.91	0.06	0.03	-0.11 to 0.45	-0.40 to 0.30
σ_2					1.00	0.90	0.06	0.03	-0.09 to 0.46	-0.35 to 0.31
β						1.00	0.07	0.03	-0.06 to 0.49	-0.52 to 0.37
γ_{y_1}							1.00	-0.32	-0.03 to 0.07	-0.07 to 0.045
γ_{y_2}								1.00	-0.04 to 0.05	-0.04 to 0.04

Table 5.4: Correlation matrix of the simulation using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors.

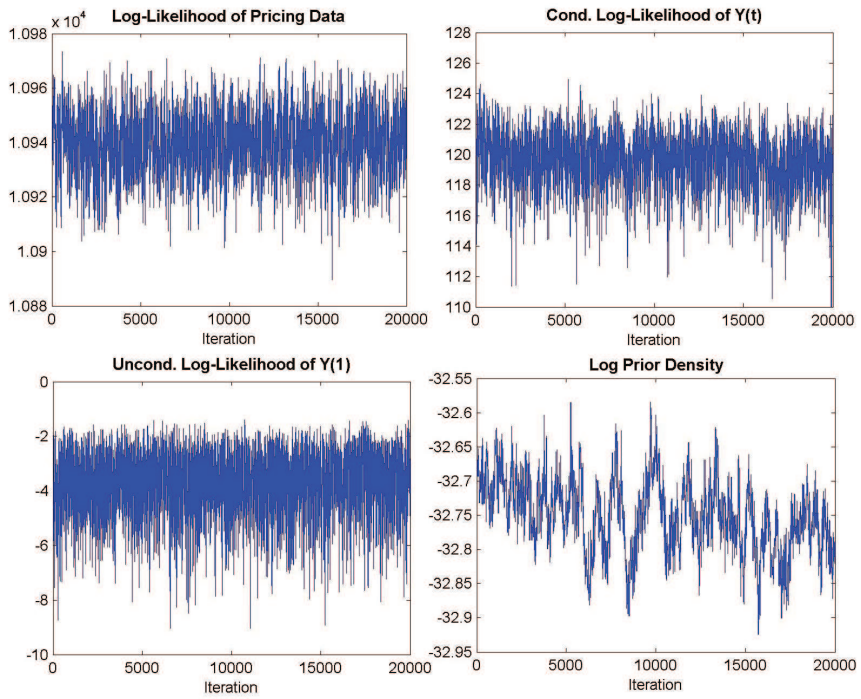

Figure 5.26: Log posterior components (referring to equation 5.9): log-likelihood of pricing data (top, left), conditional and unconditional log-likelihood of latent variables (top, right and bottom, left) and log prior density (bottom, right) using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors.

Chapter 6

Estimating Two-Factor Cairns Term Structure Models using Markov Chain Monte Carlo: UK Strips Data

From Chapter 5, we have seen that the chain convergence from estimating the two-factor Cairns term structure model using simulated data are significantly improved after we employed the adaptive Metropolis-Hastings algorithm with a blocking strategy and repameterising the bond posterior distribution. In this chapter, we therefore continue on to estimate the model using real market data.

6.1 UK Strips Data

For the market data, we consider monthly UK Strips from November 2002 to June 2008 (68 months) where the prices are taken on the last business day of each month and pooled into fixed 20 maturities: 0.25, 0.5, 1.0, 2.0, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 12.5, 15.0, 17.5, 20.0, 22.5, 25.0, 27.5, 30.0 years. Note that we simply use a linear interpolation in order to obtain the data for the same constant maturities as the simulated data in the previous chapter. The plots for UK Strips yields are shown in Figure 6.1.

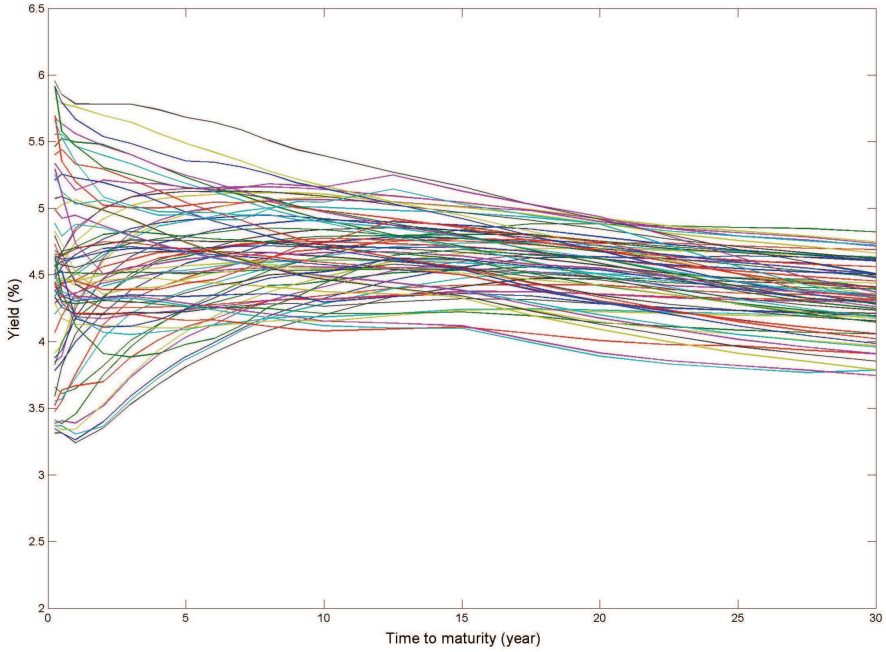

Figure 6.1: Monthly UK Strips yields from November 2002 to June 2008 (68 months).

6.2 Estimating Results on Monthly UK Strips Data

We estimate here the Cairns term structure model using monthly UK Strips data (68 months from November 2002 to June 2008 as described in Section 6.1). In general, the algorithm and procedure used are exactly the same as in Section 6.6.3 but the dataset is smaller (due to the availability of the market data). Also, the parameter σ_ε will now be estimated in terms of the precision parameter $\tau_\varepsilon = 1/\sigma_\varepsilon^2$ with prior distribution

$$f_0(\tau_\varepsilon) = \Gamma(0.01, 1.0 \times 10^8).$$

We note that the above prior has mean and standard deviation of 1.0×10^6 (equivalent to 0.001 for σ_ε) and 1.0×10^7 respectively. For the MH algorithm, τ_ε will be updated individually and a candidate point at j-th iteration will be drawn from the constant variance proposal distribution

$$q_{\tau_\varepsilon} \sim N(\tau_\varepsilon(j-1), vol_{\tau_\varepsilon}^2),$$

where vol_{τ_ε} is set equal to 750 in the simulation. It is worth mentioning that we

initially tried to use an adaptive proposal variance for τ_ε, but it turned out to be inefficient as the variance computed from previous 200 sample values tends to go to zero since the chain moves very slowly. Hence, we use a constant proposal variance for τ_ε.

In the current circumstances, we do need to run the chain much longer since the true parameter values are not known. In particular, we found that it took very long for the parameters β and ρ to start converging. After long simulations, we achieve the following results.

- We first note that all the following results are with 4,000 values by recording every 20th iteration out of 80,000 iterations. This is known as a thinning procedure which can mitigate the degree of autocorrelation in the individual simulation paths of each parameter and latent variable.

- Figure 6.2 shows the simulated chains of the model parameters and the corresponding posterior estimates are given in Table 6.1. As can be seen, all parameters converge reasonably well (though, rather slowly for parameter σ_ε). Unsurprisingly, the convergence of γ_{y_1} and γ_{y_2} is far better than for the other parameters since these two parameters appear only in the likelihood of the latent variables (not in part of the pricing formula). With the UK market data, we can observe that the estimated α_1 turns to be relatively low (mean of $\alpha_1 = 0.113$) and the latent variables Y_1 and Y_2 are strongly negatively correlated (mean of $\rho = -0.807$). Moreover, comparing these to the results obtained with the simulated data (Table 6.3), the coefficient of variation (standard deviation/mean) of almost all parameters (except ρ, γ_{y_1} and γ_{y_2}) is found to be higher, especially of β (approximately 10 times). One possible reason is that the estimated value of σ_ε on the market data now turns to be 0.0024, while it was fixed at 0.001 for the simulated dataset. Moreover, other reason is that the real data is more complicated and also shorter than the simulated data.

- Figure 6.3 illustrates the plots of the posterior densities for the model parameters. We can see that most tend to have some skewness. Regarding autocorrelation functions, it can be noticed from Figure 6.4 that there is no evidence of high autocorrelations for γ_{y_1} and γ_{y_2} while those of σ_ε are most persistent.

Clearly, we may need to run the simulation even longer and use a more extensive thinning procedure in order to improve autocorrelations, but it is not expected that the convergence and estimation will be improved since all the chains generally converged reasonably well.

- Sample paths of the latent variables $Y_1(t)$ and $Y_2(t)$, for $t = 1, 20, 30, 40, 50$, and 68, are demonstrated in Figure 6.5. We can see that the movements of each pair are likely to be in opposite direction. This is consistent with the estimating result for ρ in which the mean is about -0.807. Figure 6.6 shows plots of 95% credible intervals constructed from the sample paths with the mean values of $Y(t)$ for all t. The MH acceptance rates are between around 6% to 8%. It is obvious that the intervals in the figure are much wider than those for the estimation with the simulated data (Figure 6.21), reflecting the higher level of uncertainty of the estimation.

- In Figure 6.7, we can see that the shapes of the posterior densities for the latent variables look smoother than those for the model parameters (Figure 6.3). According to the autocorrelation functions of the latent variables (Figure 6.7), it can be observed that those of $Y_1(t)$ are very similar to those of $Y_2(t)$ in which they are slow to decline within 500 lags.

- Figure 6.9 provides the scatter plots of pairs of the model parameters with the corresponding correlation matrix given in Table 6.2. As can be seen, the high positive correlations among α_2, σ_2 and β still exist (as with the simulated data), and also σ_1 becomes more correlated with other parameters. Furthermore, it turns out that γ_{y_1} is strongly negatively correlated to γ_{y_2}, whereas the additional parameter σ_ε does not appear to be correlated with any other parameters.

- The movements of each log posterior component along the simulation are shown in Figure 6.10. Overall, all the components, especially the conditional log-likelihood of latent variables, are less stable than those with the simulated data.

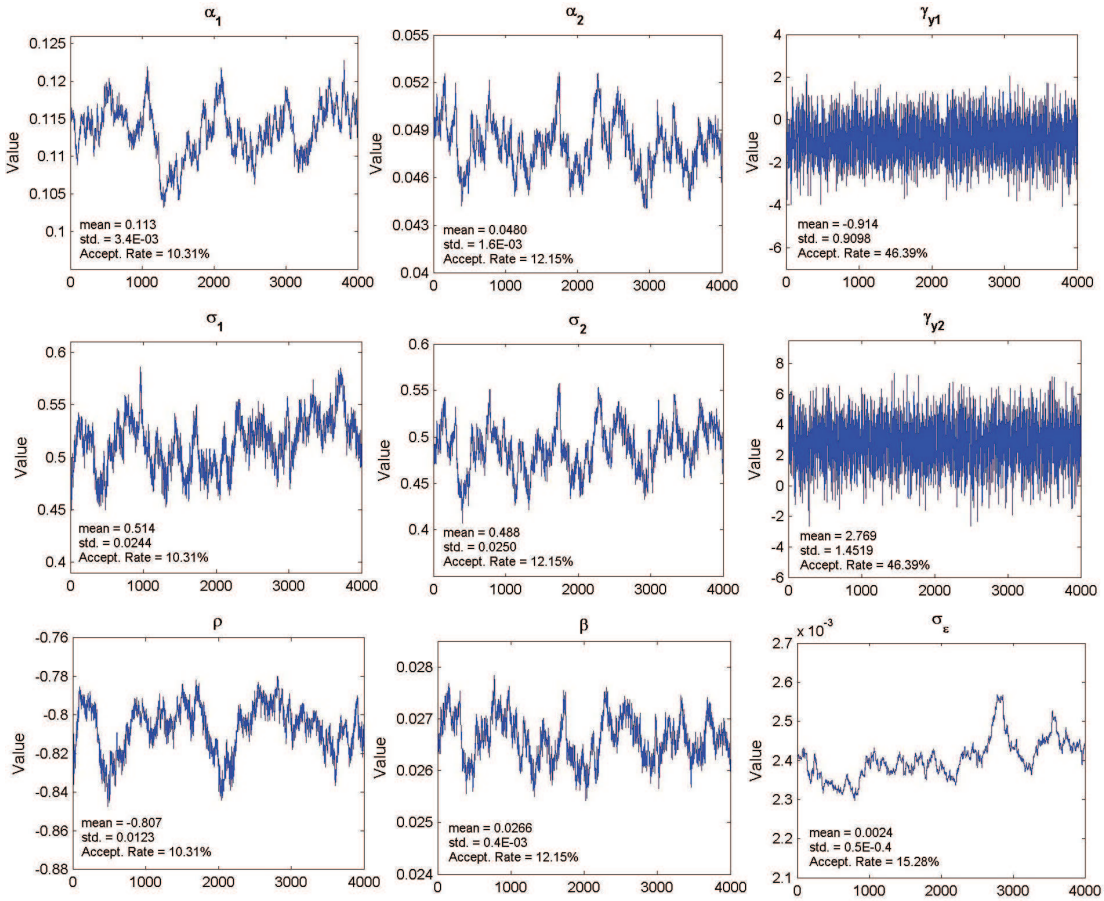

Figure 6.2: Sample paths of model parameters of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. Plots are of 4,000 values (every 20th iteration out of 80,000 iterations).

	Mean	Std.	95% Credible Interval	Acceptance rate	Scaling of the proposal std.
α_1	0.113	0.0034	(0.1055, 0.1195)	10.31%	2.0
σ_1	0.514	0.0244	(0.4682, 0.5615)	10.31%	2.0
ρ	-0.807	0.0123	(-0.8347, -0.7876)	10.31%	2.0
α_2	0.0480	0.00157	(0.04514, 0.05133)	12.15%	1.8
σ_2	0.488	0.0250	(0.4385, 0.4385)	12.15%	1.8
β	0.0266	0.00043	(0.02582, 0.02742)	12.15%	1.8
γ_{y_1}	-0.914	0.9098	(-2.6730, 0.9299)	46.39%	1.0
γ_{y_2}	2.769	1.4519	(-0.1392, 5.6265)	46.39%	1.0
σ_ε	0.0024	0.00005	(0.00232, 0.00254)	15.28%	1.0

Table 6.1: Summary statistics of parameter posterior estimates of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. The inference is made from 4,000 values (every 20th iteration out of 80,000 iterations).

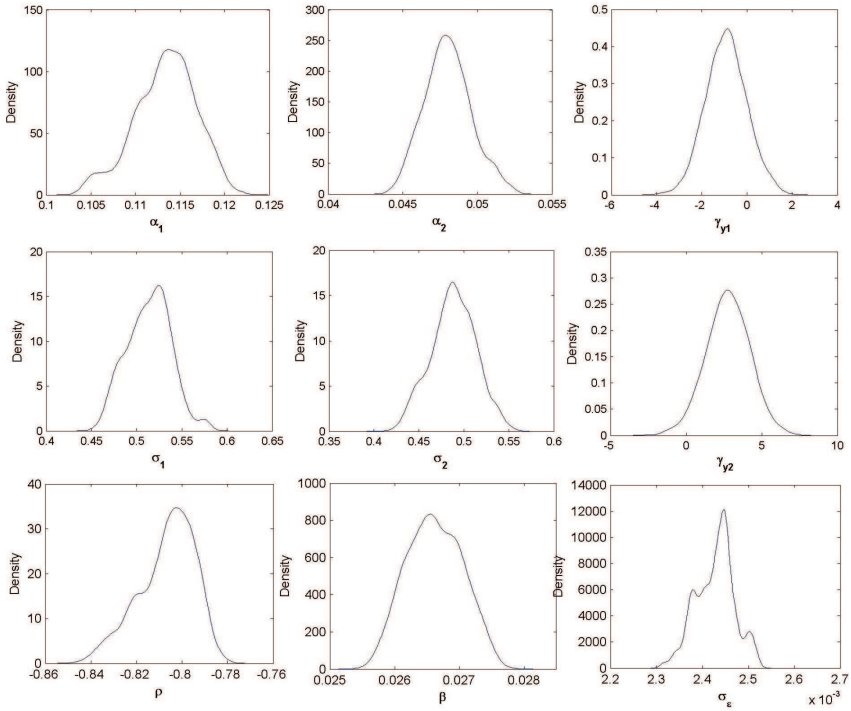

Figure 6.3: Posterior densities of model parameters of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. Plots are of 4,000 values (every 20th iteration out of 80,000 iterations).

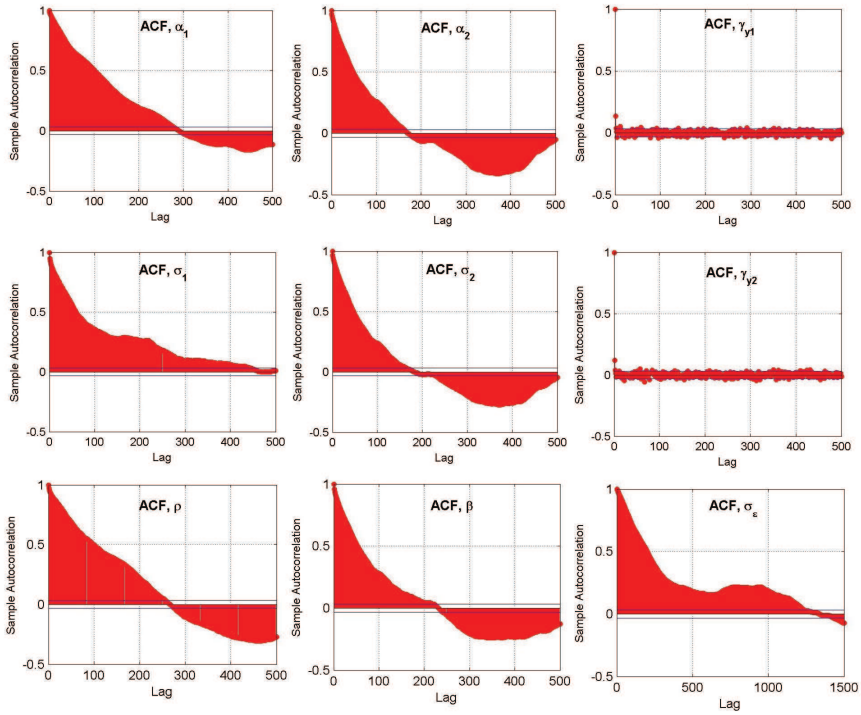

Figure 6.4: Autocorrelation functions of model parameters of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. Plots are of 4,000 values (every 20th iteration out of 80,000 iterations).

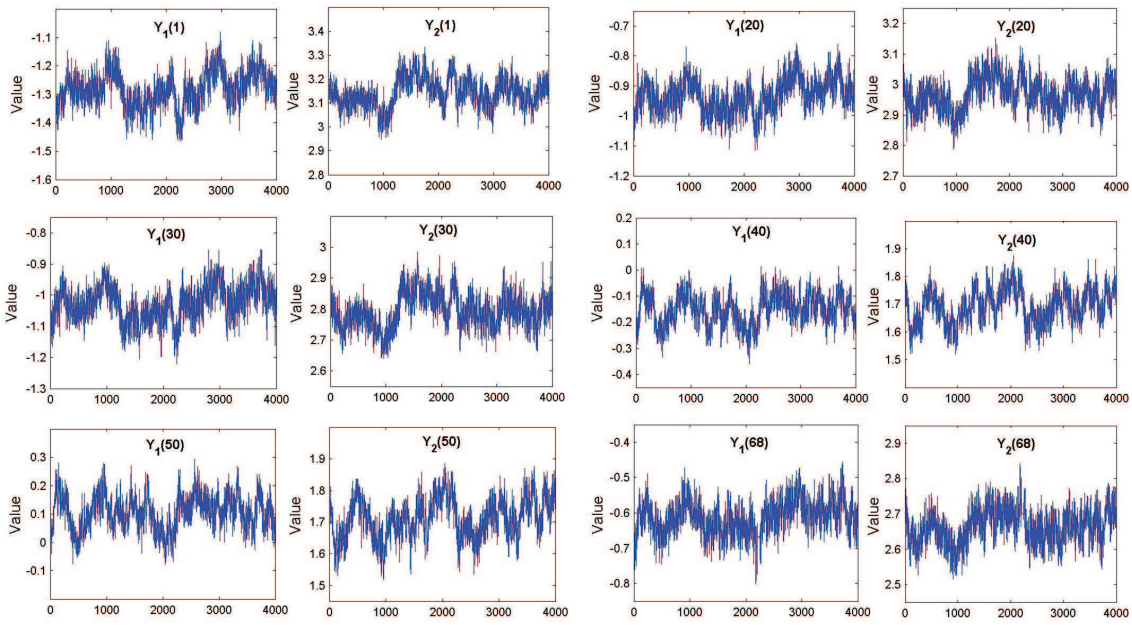

Figure 6.5: Sample paths of latent variables (for $t = 1, 20, 30, 40, 50$ and 68) of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. Plots are of 4,000 values (every 20th iteration out of 80,000 iterations).

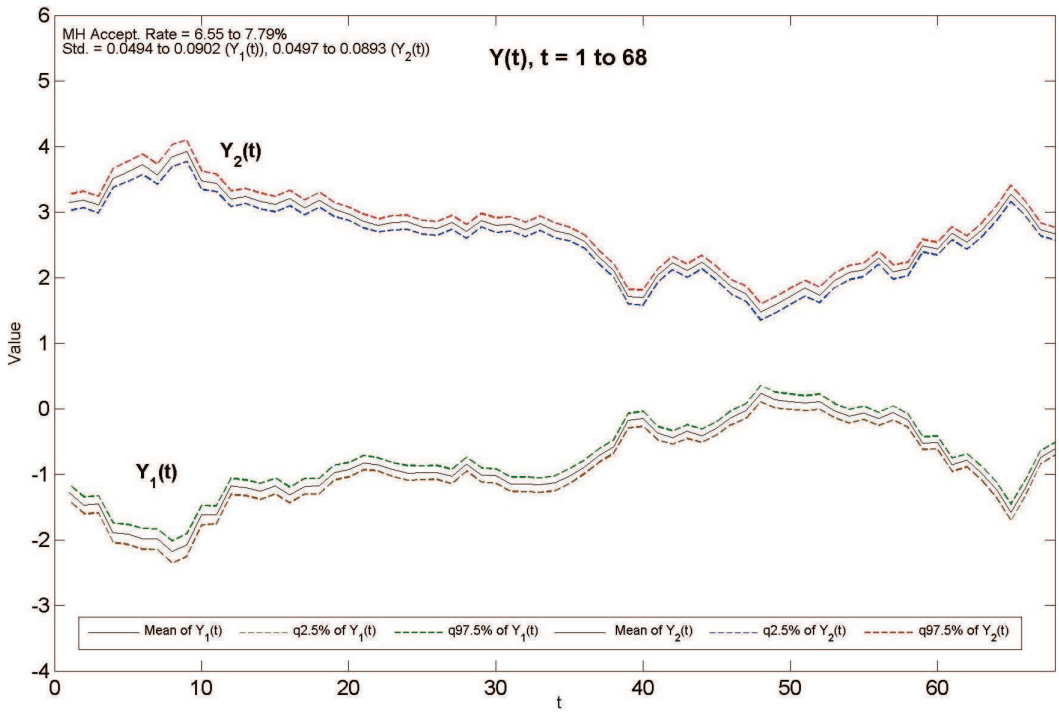

Figure 6.6: Plots of 95% credible interval constructed from the sample paths with the mean values of $Y_1(t)$ and $Y_2(t)$ for $t = 1, \ldots, 68$, of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. The inference is made from 4,000 values (every 20th iteration out of 80,000 iterations).

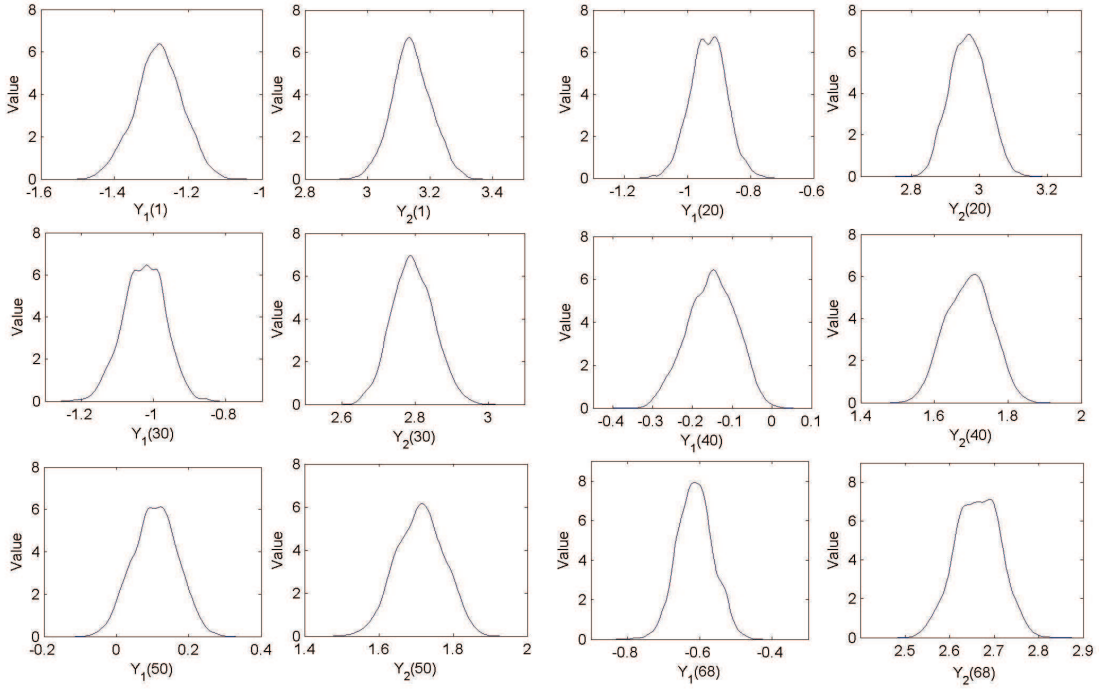

Figure 6.7: Posterior densities of latent variables (for $t = 1, 20, 30, 40, 50$ and 68) of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. Plots are of 4,000 values (every 20th iteration out of 80,000 iterations).

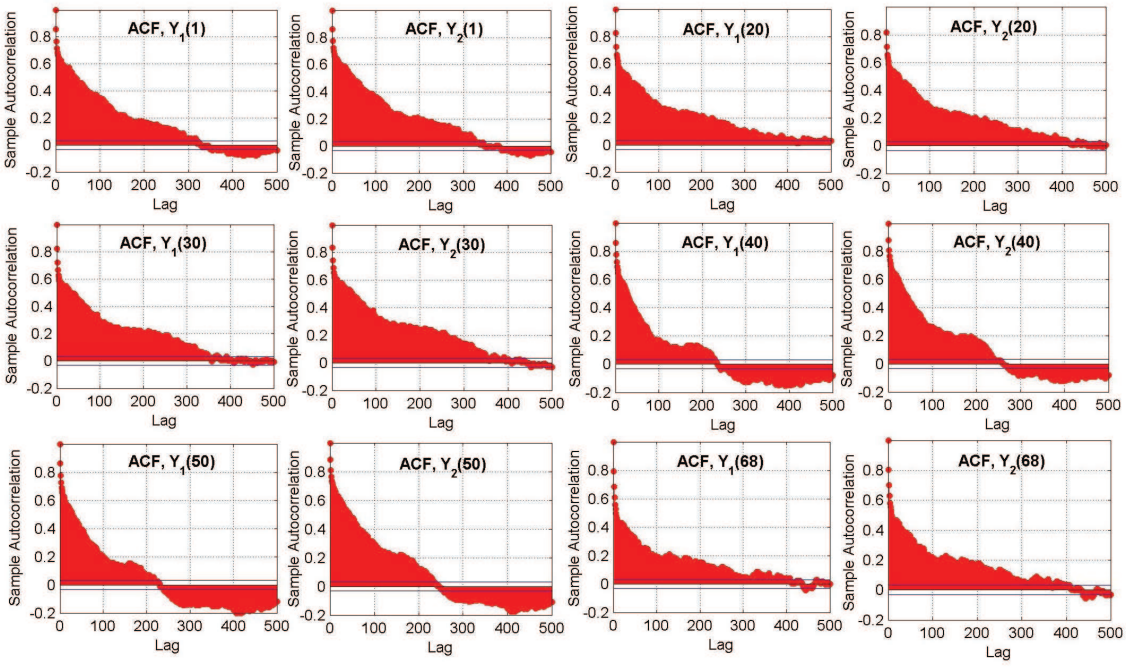

Figure 6.8: Autocorrelation functions of latent variables (for $t = 1, 20, 30, 40, 50$ and 68) of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. Plots are of 4,000 values (every 20th iteration out of 80,000 iterations).

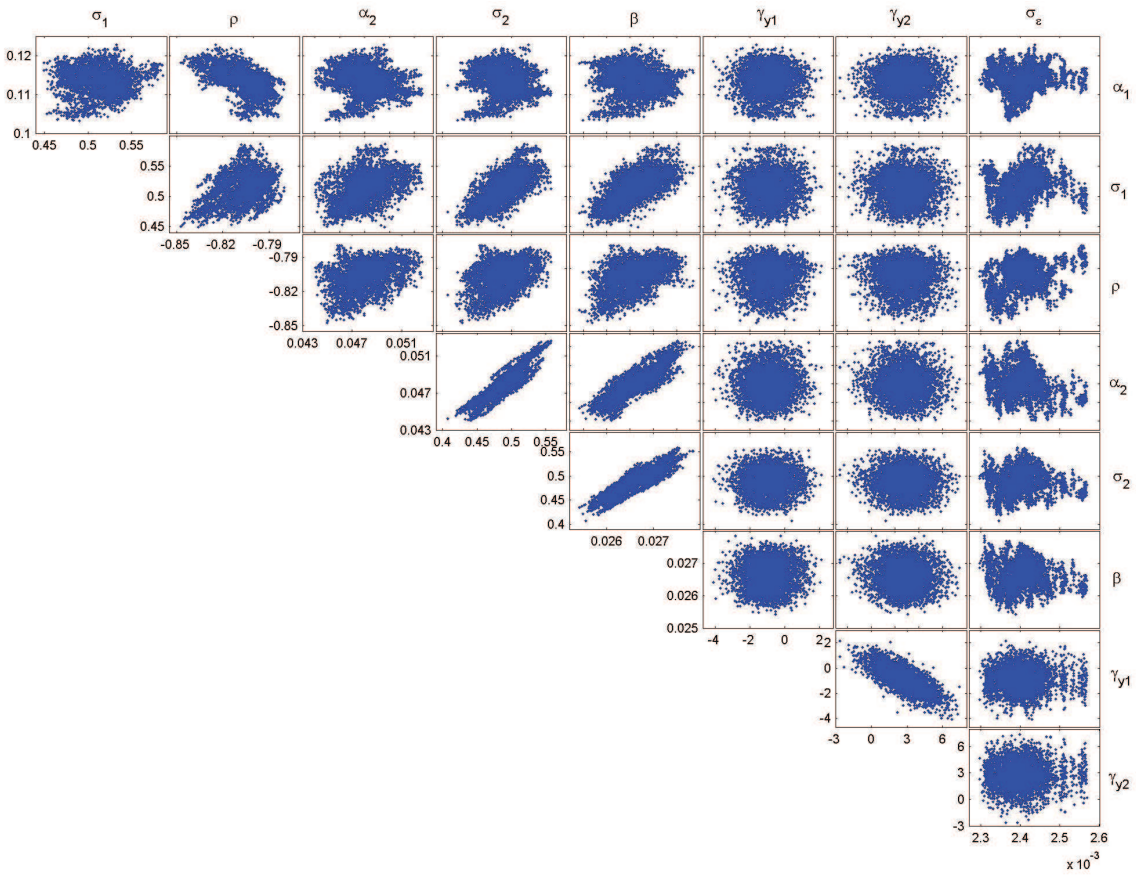

Figure 6.9: Scatter plots of model parameters of the two-factor Cairns term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. Plots are of 4,000 values (every 20th iteration out of 80,000 iterations).

	α_1	σ_1	ρ	α_2	σ_2	β	γ_{y_1}	γ_{y_2}	σ_ε	$Y_1(t)$	$Y_2(t)$
α_1	1.00	0.04	-0.58	-0.14	-0.06	-0.06	0.01	0.01	0.08	-0.39 to 0.68	-0.76 to 0.47
σ_1		1.00	0.38	0.43	0.69	0.67	0.04	-0.03	0.24	0.06 to 0.64	-0.49 to 0.25
ρ			1.00	0.39	0.39	0.48	0.03	-0.02	0.31	-0.32 to 0.71	-0.64 to 0.39
α_2				1.00	0.90	0.85	0.01	-0.04	-0.17	-0.59 to 0.56	-0.56 to 0.44
σ_2					1.00	0.87	0.02	-0.03	0.07	-0.40 to 0.63	-0.50 to 0.36
β						1.00	0.04	-0.04	-0.04	-0.28 to 0.71	-0.73 to 0.15
γ_{y_1}							1.00	-0.74	0.03	0.02 to 0.08	-0.07 to 0.05
γ_{y_2}								1.00	0.02	-0.06 to 0.03	-0.02 to 0.06
σ_ε									1.00	0.16 to 0.39	-0.14 to 0.33

Table 6.2: Correlation matrix of the simulation using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. The inference is made from 4,000 values (every 20th iteration out of 80,000 iterations).

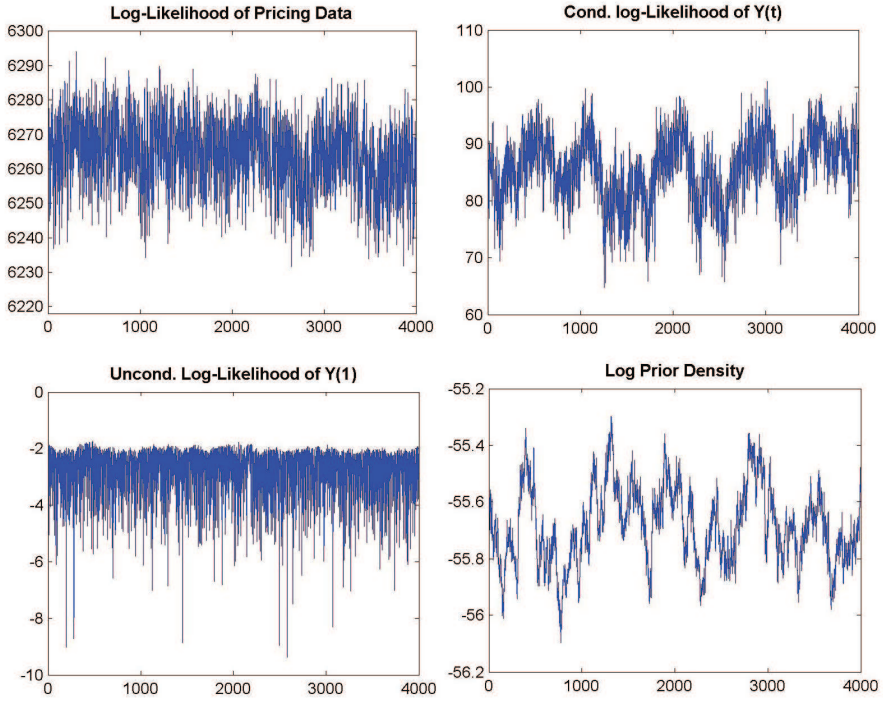

Figure 6.10: Log posterior components (referring to equation 5.9): log-likelihood of pricing data (top, left), conditional and unconditional log-likelihood of latent variables (top, right and bottom, left) and log prior density (bottom, right) using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. Plots are of 4,000 values (every 20th iteration out of 80,000 iterations).

6.3 Goodness of Fit

At the end of this chapter, we investigate the goodness of fit of the two-factor Cairns term structure model by considering bond price residuals or the differences between the market UK Strips and the estimated Cairns bond prices. Let $P(t, \tau_{tj})$ and $C(\tau_{tj}; \bar{Y}(t), \bar{\theta})$ be respectively market and estimated bond prices at time t maturing at $t + \tau_{tj}$, where $\bar{Y}(t)$ and $\bar{\theta}$ are the mean values of the latent variables and model parameters from the MCMC results. Hence, the residuals can be defined by

$$\hat{\varepsilon}(t, \tau_{tj}) = P(t, \tau_{tj}) - C(\tau_{tj}; \bar{Y}(t), \bar{\theta}). \tag{6.1}$$

Recall that we assume $\hat{\varepsilon}(t, \tau_{tj}) \sim$ i.i.d. $N(0, \sigma_\varepsilon^2)$ for each time t according to the framework in Section 5.1.

Figure 6.11 shows the bond price residual surface from November 2002 to June 2008. From the figure, we generally cannot observe any particular pattern of the residuals except that there is graphical evidence that the residuals are not independent. However, the residuals are relatively higher for the last 8 months which is the same period as the Northern Rock bank crisis started and the financial credit crisis loomed. In Figure 6.12, it can be noticed that the means of the residuals (left) are fairly in line with the model assumption that they are assumed to be zero on average (even though there is serial correlation with a trough and peaks at around the middle and the two ends respectively). For the standard deviation (right), it generally can remain the values within about $\hat{\sigma_\varepsilon}$ (the estimated value of $\sigma_\varepsilon = 0.0024$) for the first 60 months, and then dramatically increases from around the 61st month (the end of 2007) onwards.

The bond price residuals of 3-month, 5-year and 30-year maturities are particularly considered in Figure 6.13. It can be found that in all cases the bond residuals of 3-month maturity can have values in the ranges of $\pm 1.0 \times \hat{\sigma_\varepsilon}$. For the 5-year maturity, the bond residuals are out of the range for the last 5 months while the residuals of 30-year maturity are relatively higher than those of the other two maturities. We may infer from the figure that the two-factor Cairns model can fairly capture the dynamics of UK Strips prices for all three maturities during the normal market condition but poorly in the volatile market period, especially for the medium- and long-term maturities.

Figure 6.14 demonstrates normal QQ-plots of the bond price residuals for the selected months from November 2002 to June 2008. As can be seen, the distribution of the residuals may not be normal for some t. Also, most are correlated with one another over time (the correlation matrix not shown here). These suggest that our assumptions that the residuals are independent and normally distributed may not be always valid.

Next, we compare the market UK Strips to the fitted spot rate curves for the selected months from November 2002 to June 2008. In Figures 6.15a to 6.15c, black solid lines represent the fitted spot rates using the means of parameter and latent

variable values, whereas green bands are the fan charts constructed from the fitted spot rates using all 4,000 sets of parameters and latent variables from the MCMC output (the outer limits of a fan chart are the 5% and 95% quantile range). The yields are converted from the interpolated UK Strips and the fitted bond prices. From the figures, we can find that if the yield curves are in a rather simple shape (e.g. October 2004), the two-factor Cairns model fits the data reasonably well but for the more complex shapes (e.g. July 2005, March 2008 and so on), the fitting is rather poor. Particularly, in many cases, the two-factor model is unlikely to produce a steeply humped or a kink shape at the short-end of the yield curves. Also, during the volatile market period (Figure 6.15c), we can observe that the model fitting is problematic. Clearly, the time-homogeneous two-factor Cairns model cannot generate several humps in a yield curve. According to the fan charts, we found that parameter uncertainty affects the fitting of the model in many cases. For instance, in October 2004, several market yields are out of the black line but they can still remain in the green fan.

We make a final remark that for a single date of the yield curve, with suitable parameter values, the two-factor Cairns model can produce one steeply humped shape (Appendix C). However, in this case since we estimate the parameters for a whole bond price surface and the data are complex (very high variation at the short-end, particular for the last 8 months), there is a trade-off of the estimated parameter values so that they cannot be fitted as well as expected in some months. Moreover, we note that here we use the mean values of estimated parameters and latent variables from the MCMC simulation results in which we observed that the parameter uncertainty can have an impact on the fitting to an extent.

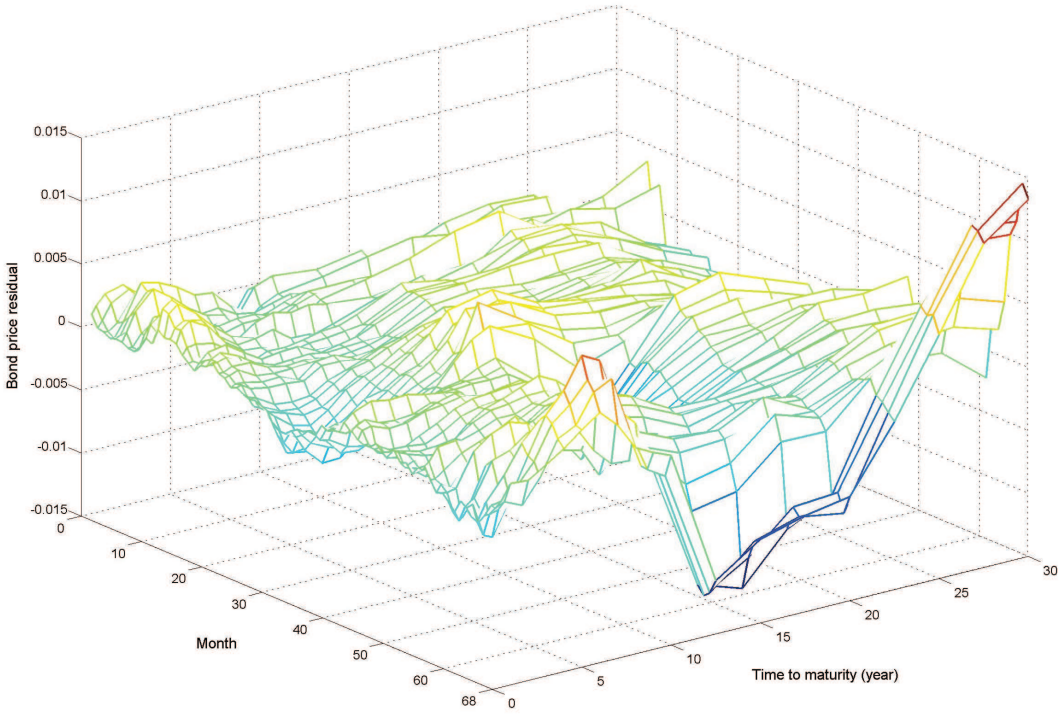

Figure 6.11: Bond price residual surface of the two-factor Cairns term structure model fitted with monthly UK Strips data from November 2002 to June 2008 (68 months).

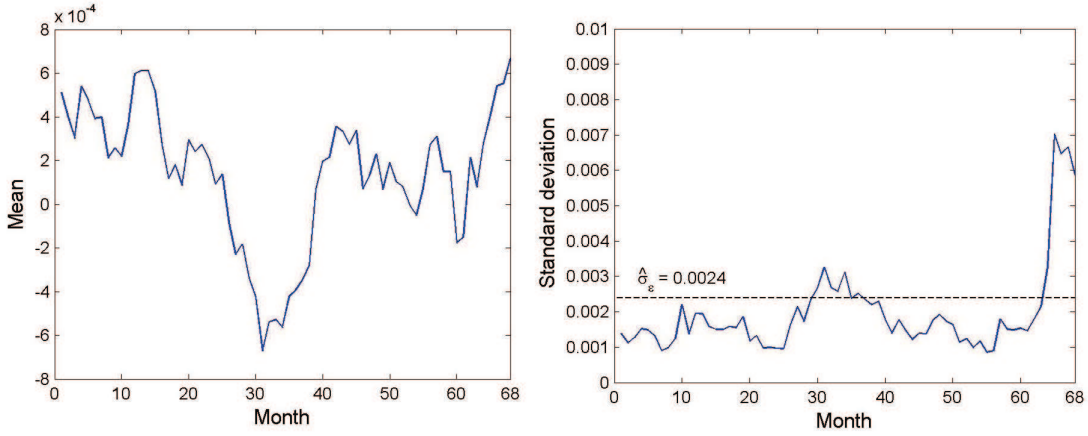

Figure 6.12: Means (left) and standard deviations (right) of the bond price residuals of the two-factor Cairns term structure model fitted with monthly UK Strips data from November 2002 to June 2008 (68 months). Dash line: the estimated $\hat{\sigma}_\varepsilon = 0.0024$.

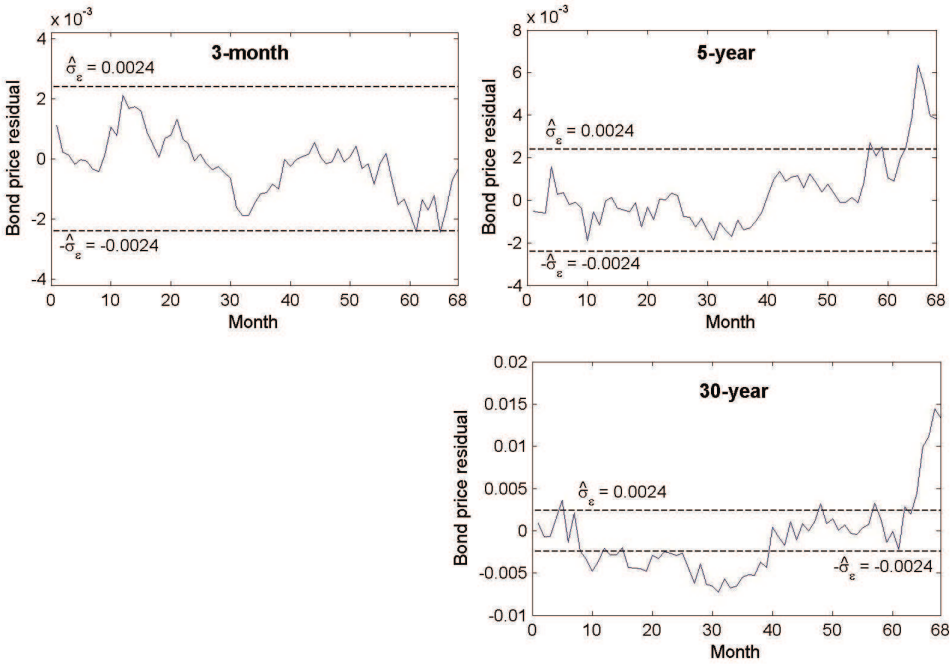

Figure 6.13: Bond price residuals of 3-month (top left), 5-year (top right) and 30-year (bottom right) maturities of the two-factor Cairns term structure model fitted with monthly UK Strips data from November 2002 to June 2008 (68 months). Dash lines: the intervals of $\pm 1.0 \times \hat{\sigma}_\varepsilon$, where $\hat{\sigma}_\varepsilon = 0.0024$.

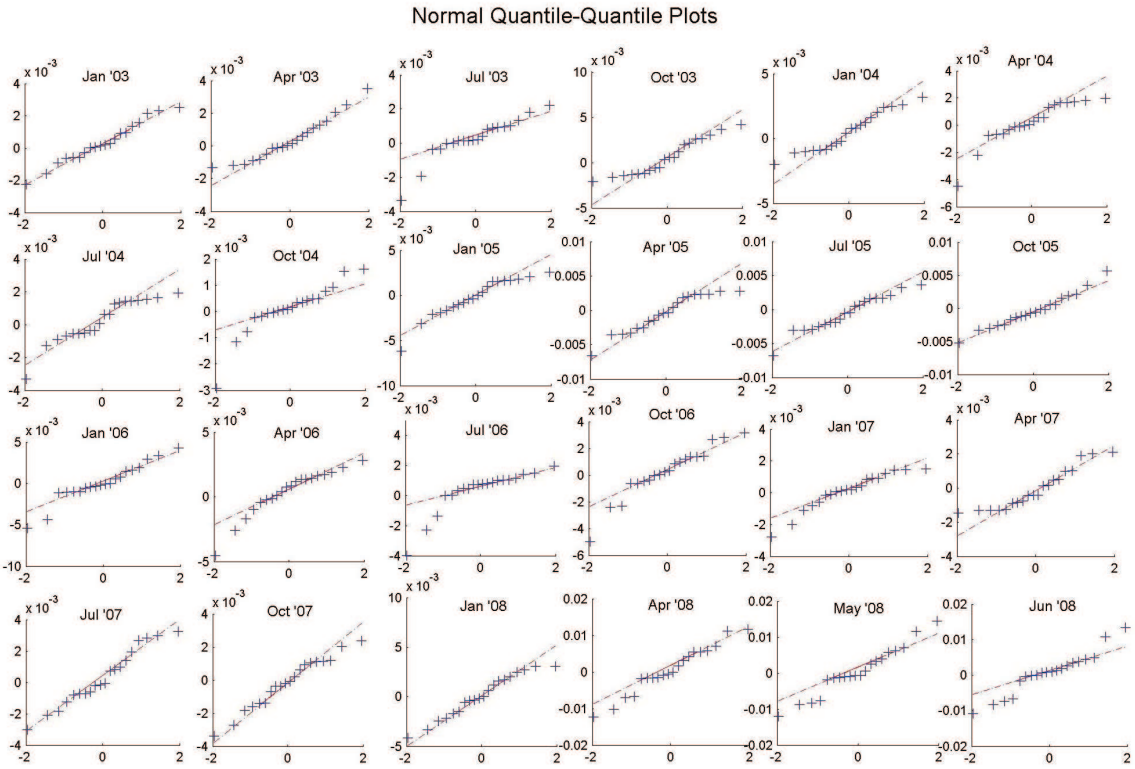

Figure 6.14: Normal QQ-plots of bond price residuals of the two-factor Cairns term structure model fitted with monthly UK Strips data for the selected months from November 2002 to June 2008.

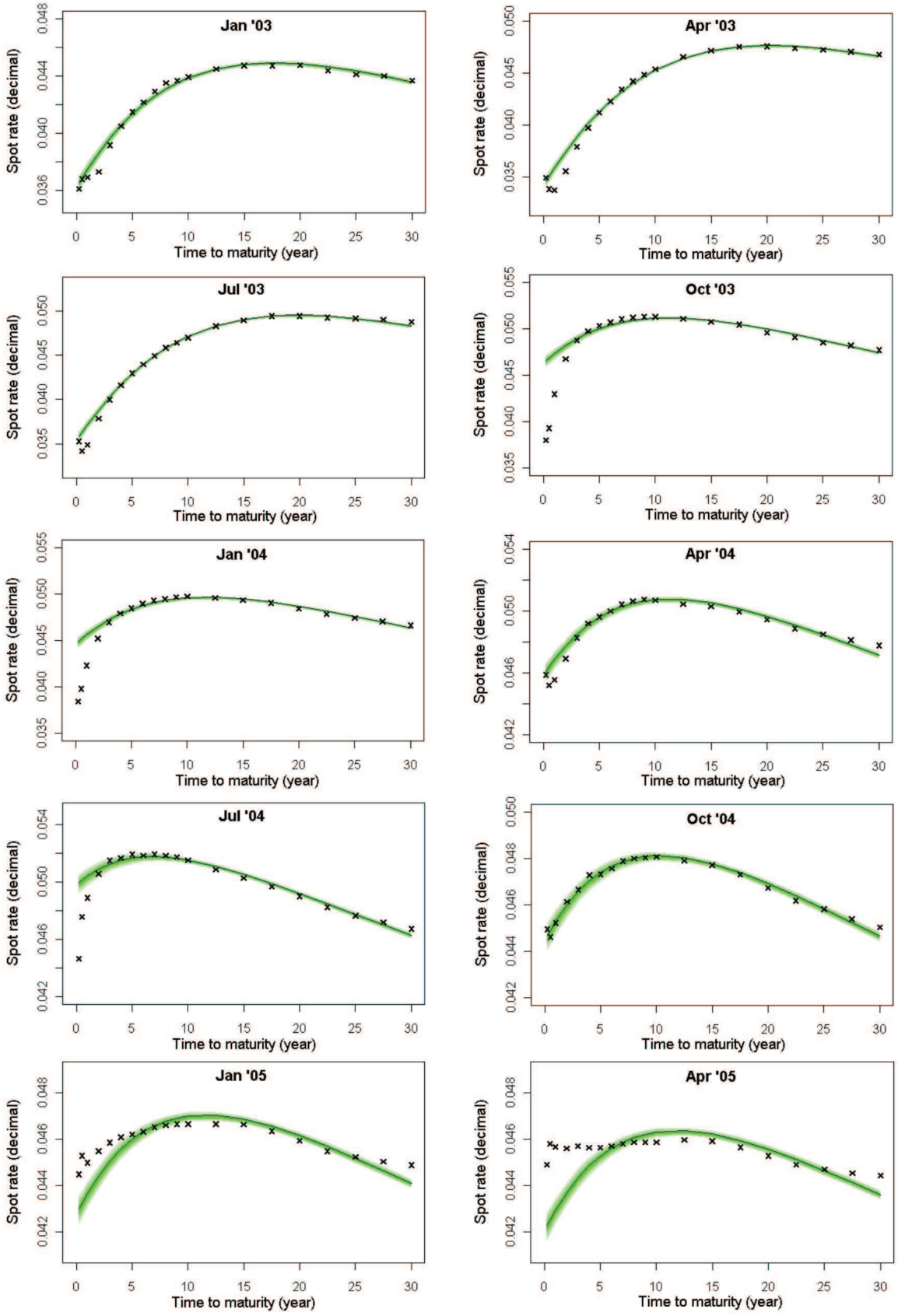

Figure 6.15a: UK Strips yields (cross mark) compared with the fitted spot rates (solid) of the two-factor Cairns term structure model using the means of parameters and latent variables for the selected months from November 2002 to April 2005. Green bands: fan charts constructed from the the fitted spot rates using all 4,000 sets of parameters and latent variables from the MCMC output.

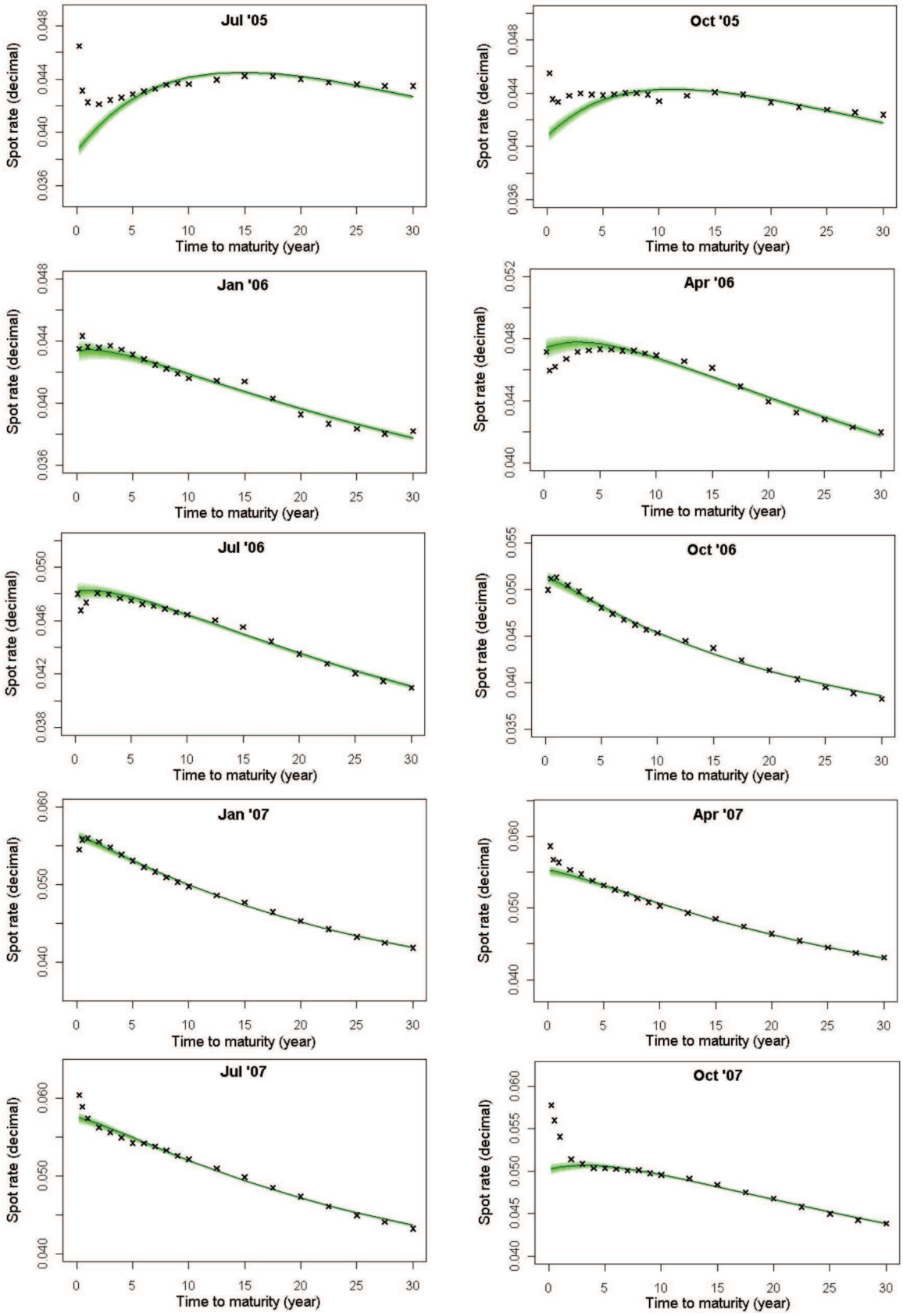

Figure 6.15b: UK Strips yields (cross mark) compared with the fitted spot rates (solid) of the two-factor Cairns term structure model using the means of parameters and latent variables for the selected months from May 2005 to October 2007. Green bands: fan charts constructed from the the fitted spot rates using all 4,000 sets of parameters and latent variables from the MCMC output.

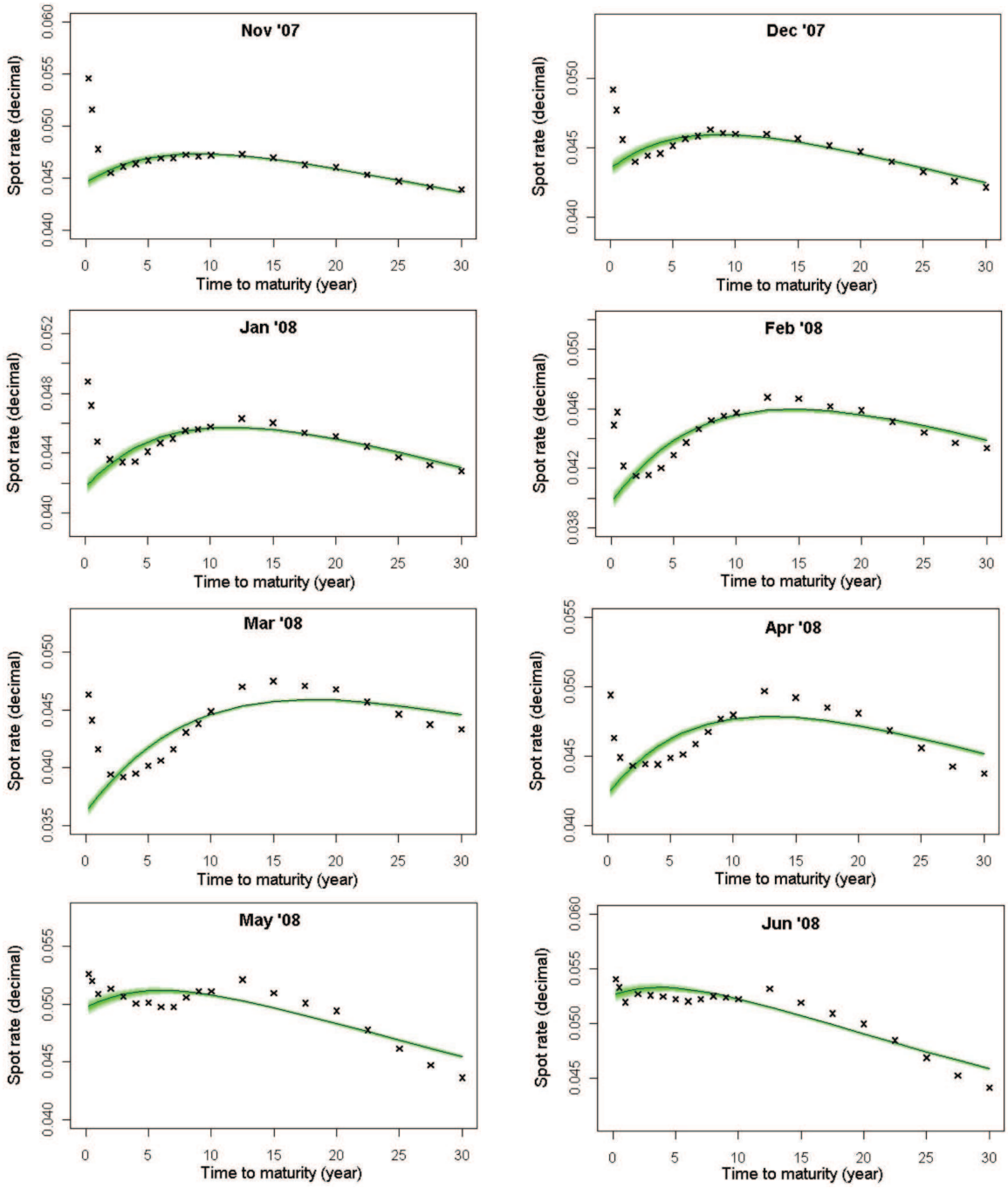

Figure 6.15c: UK Strips yields (cross mark) compared with the fitted spot rates (solid) of the two-factor Cairns term structure model using the means of parameters and latent variables for the last 8 months from November 2007 to June 2008. Green bands: fan charts constructed from the the fitted spot rates using all 4,000 sets of parameters and latent variables from the MCMC output.

Chapter 7

Forecasting Yield Curves with Parameter Uncertainty: An Application of MCMC

In this chapter, we assess the impact of parameter uncertainty on the forecasting of yield curves using the MCMC output from Chapter 6. To begin with, the main sources of uncertainty that may arise from the modelling are roughly presented. Next, we define the predictive density of the Cairns bond prices and then describe the forecasting simulation procedure. In the end, the results of forecasts with parameter uncertainty will eventually be discussed.

7.1 Introduction

Uncertainty naturally occurs in most estimation problems even if a good model and technique are being used. With reference to Cairns (2000), uncertainty may arise from three main sources: process, model and parameters. Process risk is meant to be the randomness inherent in the underlying structural stochastic process. Model uncertainty refers to the choice of model in situations where the true model is unknown. What is of particular interest in this chapter is parameter uncertainty, a risk that is often ignored when a model is implemented.

By parameter uncertainty, we typically mean the uncertainty in the parameter values in a selected model. Given the availability of large data, we can still never know the parameter values with certainty. For example, the maximum likelihood method

provides us the parameter values that are "most likely" from the data. Accordingly, the impact of parameter uncertainty should be taken into an account for use of any model, particularly in a long-term horizon where uncertainty is generally magnified.

In Bayesian paradigm, parameters are treated as random variables so that it explicitly gives us a coherent framework to quantify the additional impact of parameter uncertainty on the particular financial quantities that we are interested in. In the following section, we first consider the predictive density of the Cairns bond prices.

7.2 Forecasting Cairns Bond Prices

Denote $\mathbb{P}_M$ as all historical bond prices from time $t_1, \ldots, t_M$. Given the reparameterised bond posterior distribution in (5.21), the h-year ahead conditional predictive density of the Cairns bond price for the maturities τ_j, for $j = 1, \ldots, N$, can be defined by

$$\begin{aligned} f(P(t_M + h, \tau_j)|\mathbb{P}_M, \Theta(t_M)) &= \int\int f_1(P(t_M + h, \tau_j)|Y(t_M + h), \theta) \\ &\quad \times f_{2c}(Y(t_M + h)|Y(t_M), \theta_2) \\ &\quad \times f_{2u}(Y(t_M)|\theta, \mathbb{P}_M) \times f_0(\theta) dY(t_M + h) d\theta, \end{aligned} \tag{7.1}$$

where $\Theta(t_M) = (Y(t_M), \theta)$, f_1 is the normal density function of the bond prices, f_{2c} and f_{2u} are respectively the conditional and unconditional densities of the latent variables, and f_0 is the prior.

7.2.1 Forecasting Bond Prices with Parameter Uncertainty

Having obtained the MCMC output from previous chapter, simulating the forecast bond price $P(t_M + h, \tau_j)$ according to (7.1) is straightforward. Specifically, the sample values of $\Theta(t_M)$ for all iterations can be used to simulate the latent variable $Y(t_M + h)$ and hence compute the bond price $P(t_M + h, \tau_j)$ (which inherently includes the effect of parameter uncertainty). Initially, we present a rough procedure for simulating the forecasting of quantity of interest as follows.

1. Simulate Θ from the posterior distribution (already completed in Chapter 6).

2. Extract $Y(t_M), \theta$ from Θ.

3. Simulate $Y(t_M + h)$ given $Y(t_M), \theta$.

4. Calculate quantity of interest at $t_M + h$ given the values of $Y(t_M + h)$ and θ.

The detailed procedure to simulate the h-year ahead forecasting Cairns bond prices is described below.

1. Select the index $k = k(i)$, for $i = 1, \ldots, I$, from K iterations of MCMC at random. Therefore, we have $\theta^{(k(i))}, Y^{(i)}(t_M)$, where $Y^{(i)}(t_M) := Y^{k(i)}(t_M)$.

2. Simulate the latent variables $Y^{(i)}(t_M + h)$ from

$$\begin{pmatrix} Y_1^{(i)}(t_M+h) \\ Y_2^{(i)}(t_M+h) \end{pmatrix} = \begin{pmatrix} \gamma_{y_1}^{(k(i))} + (Y_1^{(i)}(t_M) - \gamma_{y_1}^{(k(i))})e^{-\alpha_1^{(k(i))}h} \\ \gamma_{y_2}^{(k(i))} + (Y_2^{(i)}(t_M) - \gamma_{y_2}^{(k(i))})e^{-\alpha_2^{(k(i))}h} \end{pmatrix} + B^{(i)}A^{(i)}Z(t_M + h),$$

where $Z = (Z_1, Z_2)$ with $Z_1(t_M + h), Z_2(t_M + h) \sim i.i.d.\ N(0, 1)$ and

$$B^{(i)} = \begin{pmatrix} \sqrt{\sigma_{11}} & 0 \\ 0 & \sqrt{\sigma_{22}} \end{pmatrix}, \; A^{(i)} = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix},$$

$$\Rightarrow B^{(i)}C^{(i)}Z(t_M + h) = \begin{pmatrix} \sqrt{\sigma_{11}}Z_1(t_M+h) \\ \rho\sqrt{\sigma_{22}}Z_1(t_M+h) + \sqrt{1-\rho^2}\sqrt{\sigma_{22}}Z_2(t_M+h) \end{pmatrix}$$

such that the covariance matrix is

$$\Sigma^{(i)} \;=\; A^{(i)}A'^{(i)} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix},$$

where

$$\begin{aligned} \sigma_{11} &= \frac{\sigma_1^{(k(i))^2}}{2\alpha_1^{(k(i))}}(1 - e^{-2\alpha_1^{(k(i))}h}), \\ \sigma_{12} = \sigma_{21} &= \frac{\rho^{(k(i))}\sigma_1^{(k(i))}\sigma_2^{(k(i))}}{\alpha_1^{(k(i))} + \alpha_2^{(k(i))}}(1 - e^{-(\alpha_1^{(k(i))} + \alpha_2^{(k(i))})h}), \\ \sigma_{22} &= \frac{\sigma_2^{(k(i))^2}}{2\alpha_2^{(k(i))}}(1 - e^{-2\alpha_2^{(k(i))}h}) \end{aligned}$$

and hence $\rho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$.

3. Then, compute the forecast bond price for the maturities τ_j, for $j = 1, \ldots, N$

$$P^{(i)}(t_M + h, \tau_j)|\, Y^{(i)}(t_M + h), \theta^{(k(i))} = C_Y(\tau_j, Y^{(i)}(t_M + h), \theta^{(k(i))}),$$

where $C_Y(\tau_j, Y^{(i)}(t_M + h), \theta^{(k(i))})$ is the bond price by the two-factor Cairns model defined by

$$C_Y(\tau, y, \theta) = \frac{\int_\tau^\infty H(u, y)du}{\int_0^\infty H(u, y)du},$$

$$H(u, y) = \exp\Big[-\beta u + \sum_{i=1}^{2} y_i e^{-\alpha_i u} - \frac{1}{2} \sum_{i,j=1}^{2} \frac{\rho_{ij}\sigma_i\sigma_j}{\alpha_i + \alpha_j} e^{-(\alpha_i+\alpha_j)u} \Big].$$

7.2.2 Forecasting Bond Prices with Parameter Certainty

We here define "parameter certainty" as the point estimates of the MCMC output. The forecast bond prices with parameter certainty therefore can be simulated using the same procedure as with the parameter uncertainty described earlier where the means of the parameter and latent variable values will be used instead of the selected $\theta^{(k(i))}, Y^{(i)}(t_M)$ for each $k(i)$. The rough procedure is outlined below.

1. Let $\bar{\Theta}$ be the mean of the posterior distribution for Θ.
2. Extract $Y(t_M), \theta$ from $\bar{\Theta}$.
3. Simulate $Y(t_M + h)$ given $Y(t_M), \theta$.
4. Calculate quantity of interest at $t_M + h$ given the values of $Y(t_M + h)$ and θ.

7.3 Forecasting Results on Monthly UK Strips Data

To forecast the Cairns bond prices, we use the MCMC output achieved from Chapter 6 of estimating the two-factor Cairns term structure model using $M = 68$ months of UK Strips data from November 2002 to June 2008 (i.e. for each parameter and latent

variable, we have 4,000 values of every 20th iteration out of 80,000 iterations). Furthermore, the Trapezoidal rule with step size $h = 0.05$ year is used for approximating the bond prices and the forecasted bond prices will be converted to spot rates. The forecasting results are achieved as follows.

Figures 7.1, 7.2 and 7.3 show fan charts for the forecast spot rate curves with $h = 1/12, 3/12$ and 1 year where the normal random quantities Z_1 and Z_2 are fixed at 8 specific values. The outer limits of a fan chart are the 5% and 95% quantiles and the darkest band in the middle encompasses the 45% and 55% quantile range. Each fan chart are constructed by using all 4,000 sets of the parameter and latent variable values from the MCMC output, whereas the black solid lines represent the forecast spot rate curves with parameter certainty using the means of all the parameter and latent variable values. According to the results, we can observe that given the same time horizon of forecasting, different values of Z_1 and Z_2 give rise to different shapes of the forecast yield curves, particularly at the short-end in which the higher value of Z_1 tends to make the yield curve more inverted. Furthermore, it is clear that the impact of parameter uncertainty on the forecast yield curves at short-end is generally higher than the long-end. Additionally, when the time horizon of forecasting is longer (from 1/12 to 3/12 to 1 year ahead), it can be noticed that the effect of parameter uncertainty is amplified for the whole forecast yield curves, especially at the short-end.

Figures 7.4, 7.5 and 7.6 demonstrate empirical distributions of the forecast 3-month, 5-year and 30-year spot rates for 1 year ahead with parameter uncertainty (PU) and parameter certainty (PC). In each figure, there are 4,000 lines of distributions for PU case (plus an additional line for PC case) in which each line represents an empirical distribution of the forecast spot rates using one set of the parameter and latent values incorporated with "same" 500 values of future normal randomness Z_1 and Z_2. The 1-year horizon is considered in particular in order to be in line with the Solvency II regulation that mandates a one-year risk horizon at the 99.5% confidence level. From the figures, we can see that the impact of parameter uncertainty is rather high for all three forecast rates, especially on the forecast 3-month and 5-year spot rates which have wider ranges of the distributions for a given forecast rate. However, when considering the full PU case (red solid lines) by using all 4,000 $\times$ 500 values of

PU case compared with the PC case, we can observe that the uncertainty is reflected through the fatter tails of the distributions at both side for all forecast spot rates. In addition, plots of the forecast spot rates at 0.5% and 99.5% quantiles for all three maturities for PU and PC cases are also presented in Figure 7.7. As can be seen, the ranges of values at both quantiles for the 30-year maturity are much narrower than the other maturities, where the values for PC case reasonably lie in the middle of the ranges in all cases.

Next, we investigate further the most critical sets of the parameters and latent variables $\Theta^{(k)}(t_M) = (Y^{(k)}(t_M), \theta^{(k)})$ at the 0.5% and 99.5% quantiles of all three forecast rates for 1 year ahead as shown in Table 7.1. Comparing parameter and latent variable values of the critical sets to their means (Table 7.2), we can notice that generally most are not significantly different except a combination of γ_{y_1} and γ_{y_2}. Additionally, we also found that the $\Theta^{(199)}$ and $\Theta^{(3072)}$ respectively have the most and the least influence on the median of the empirical distributions for all three maturities of the forecast spot rates. We may infer from these results that the market price of risk parameters tend to have most impact on the uncertainty of the forecast spot rates for the considered 1-year time horizon.

1 year ahead forecast spot rate	$\Theta^{(k)}(t_M)$ at risk	
	0.5% quantile	99.5% quantile
3-month	$\Theta^{(3864)}$	$\Theta^{(3072)}$
5-year	$\Theta^{(199)}$	$\Theta^{(3072)}$
30-year	$\Theta^{(199)}$	$\Theta^{(3739)}$

Table 7.1: Most critical sets of the parameters and latent variables $\Theta^{(k)}(t_M)$ of the 3-month, 5-year and 30-year forecast spot rates for 1 year ahead at the 0.5% and 99.5% quantiles.

$\Theta^{(k)}(t_M)$	β	α_1	α_2	σ_1	σ_2	ρ	γ_{y_1}	γ_{y_2}	$Y_1(t_M)$	$Y_2(M)$
$\Theta^{(199)}$	0.0270	0.114	0.049	0.515	0.494	-0.796	-3.535	2.191	-0.57	2.59
$\Theta^{(3072)}$	0.0264	0.116	0.047	0.516	0.482	-0.803	2.065	1.616	-0.61	2.71
$\Theta^{(3739)}$	0.0271	0.114	0.048	0.560	0.515	-0.806	0.014	5.183	-0.53	2.63
$\Theta^{(3864)}$	0.0266	0.118	0.049	0.547	0.508	-0.805	-2.979	2.693	-0.67	2.76
$\bar{\Theta}$	0.0266	0.113	0.048	0.514	0.488	-0.807	-0.914	2.769	-0.61	2.66
Std.	0.00043	0.0034	0.0016	0.0244	0.0250	0.0123	0.9098	1.4519	0.050	0.050

Table 7.2: Parameter and latent variable values of the most critical $\Theta^{(k)}(t_M)$.

In Figures 7.8, 7.9 and 7.10, we project 3-month, 5-year and 30-year spot rates for next 10 years using fan charts. For each fan chart, we simulate the forecast spot rates for 2, 4, 6, 8 and 10 years ahead where each set of parameter and latent variable values are incorporated with 100 (for PU case, red fan) and 400,000 (for PC case, green fan) fresh pairs of future normal randomness Z_1 and Z_2. From the figures, we can see that when the forecasting time horizon is longer, the gaps between red and green fans are generally wider (especially for 3-month and 5-year forecast spot rates), reflecting the higher allowance required for parameter uncertainty. Figure 7.11 illustrates fan charts for the forecast spot rates at 10 years ahead by maturities corresponding to Figures 7.8, 7.9 and 7.10.

Figure 7.12 demonstrates scatter plots of the forecast latent variables $Y_1(t)$ and $Y_2(t)$ at 10 years ahead by parameter uncertainty (PU) and parameter certainty (PC). As expected, both have an ellipse shape which is the shape of the bivariate normal. Furthermore, the ellipse for PU case spreads over the ellipse for PC case where their means roughly remains at the same center point.

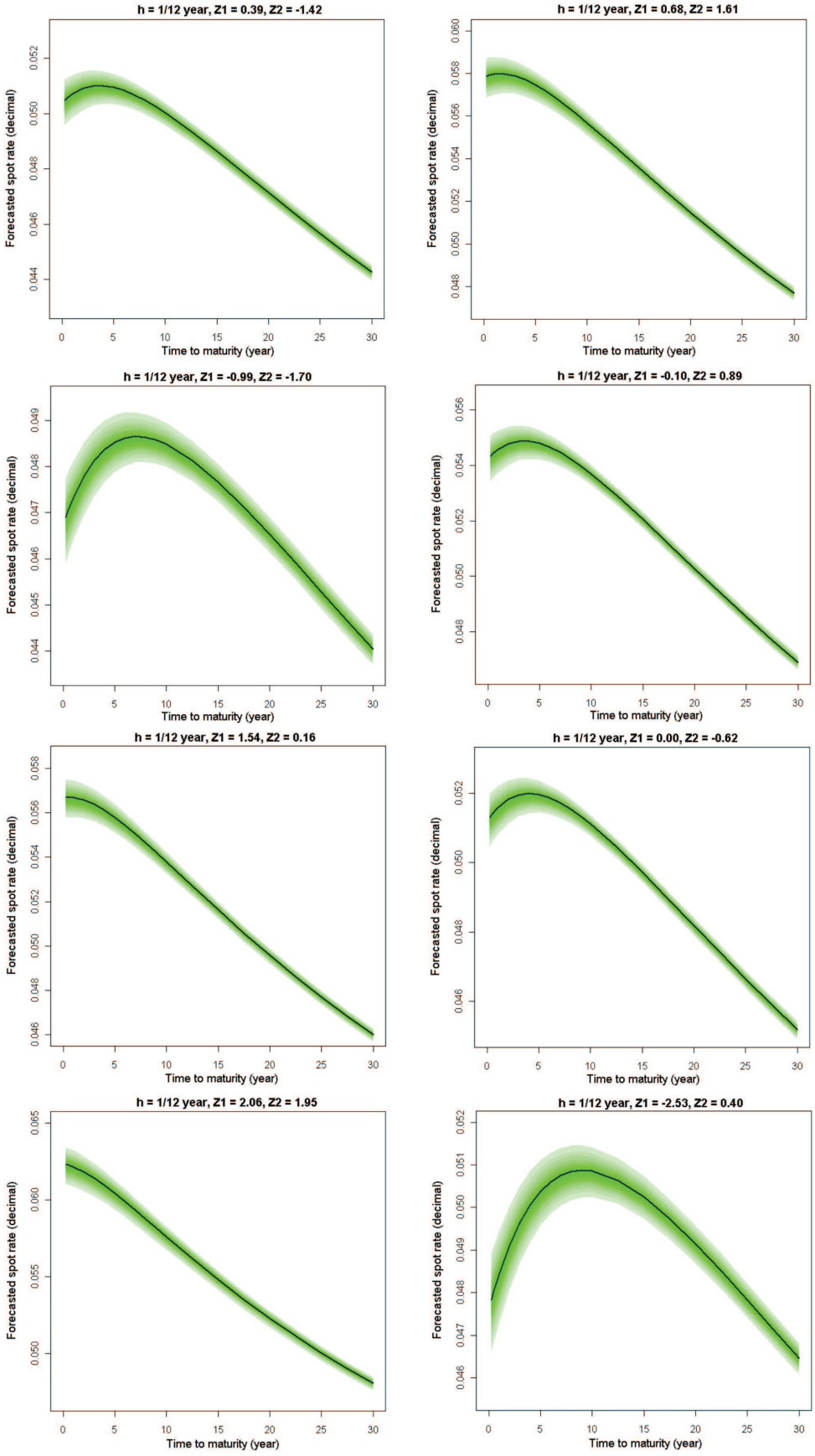

Figure 7.1: Fan charts for the forecast spot rate curves with $h = 1/12$ year by fixing Z_1 and Z_2 at 8 specific values. Each fan is constructed by using all 4,000 sets of the parameter and latent variable values from the MCMC output. Black (solid) line: the forecast yield curves with parameter certainty (PC) using the mean values of the MCMC output.

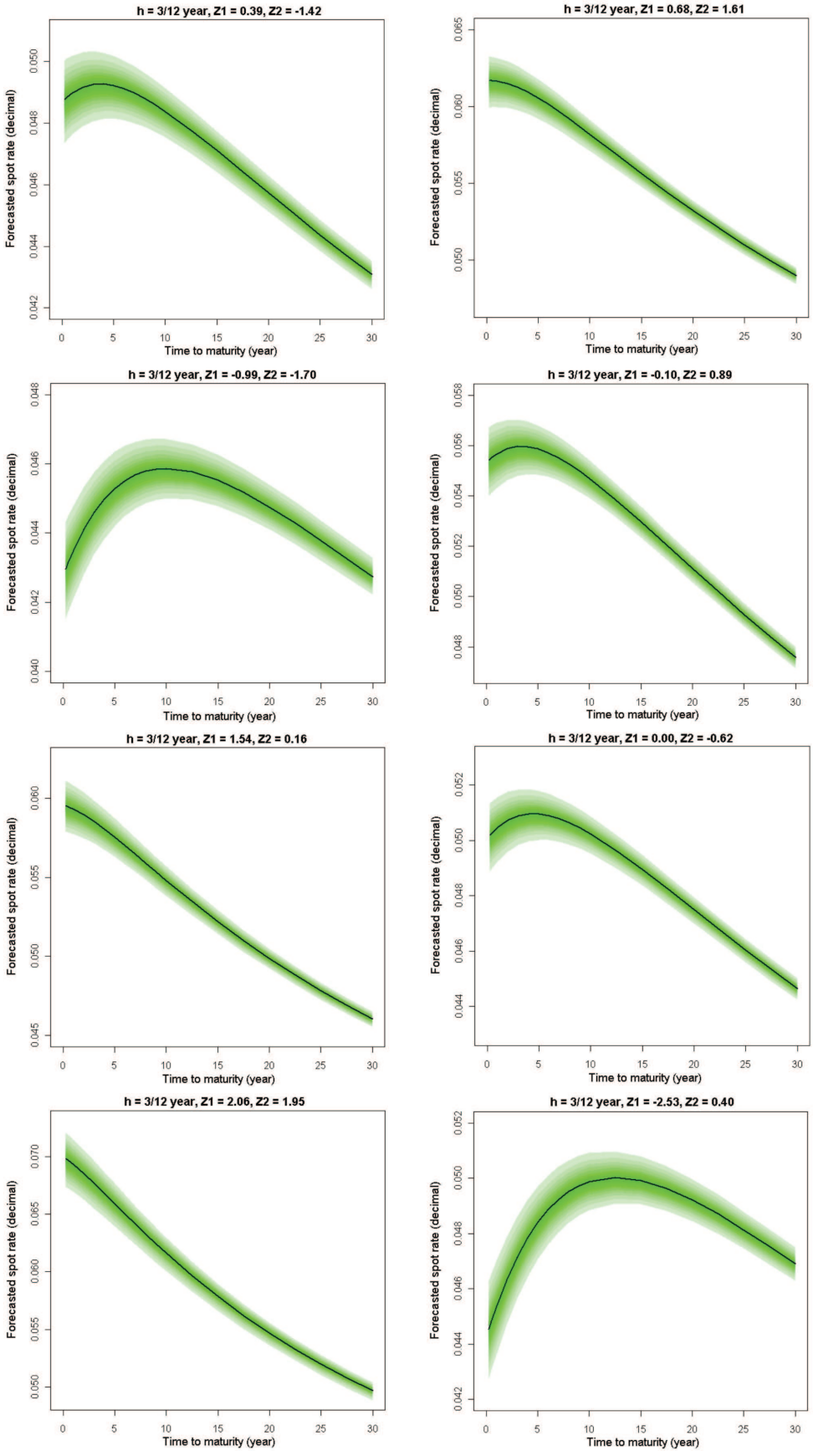

Figure 7.2: Fan charts for the forecast spot rate curves with $h = 3/12$ year by fixing Z_1 and Z_2 at 8 specific values. Each fan is constructed by using all 4,000 sets of the parameter and latent variable values from the MCMC output. Black (solid) line: the forecast yield curves with parameter certainty (PC) using the mean values of the MCMC output.

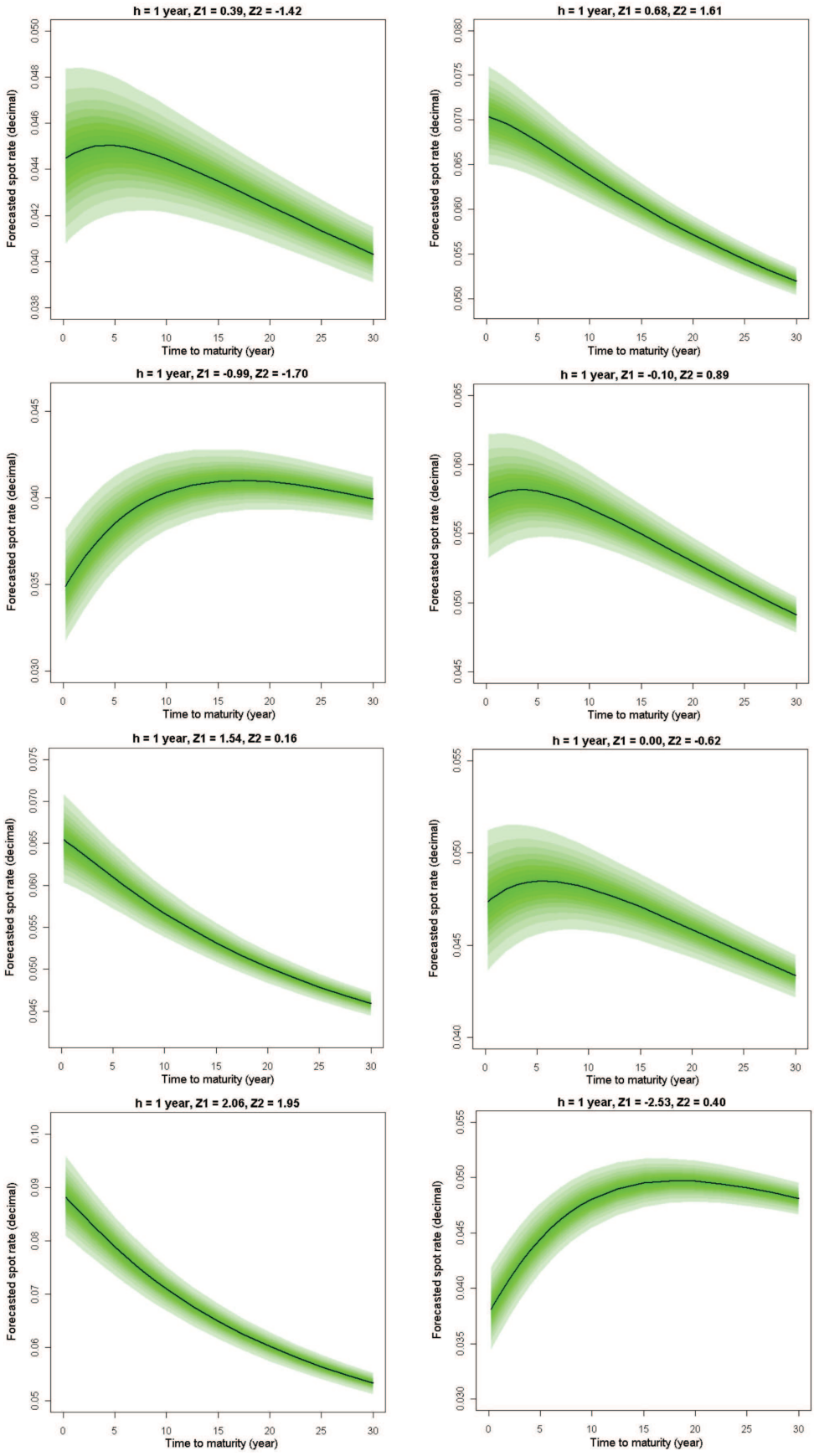

Figure 7.3: Fan charts for the forecast spot rate curves with $h = 1$ year by fixing Z_1 and Z_2 at 8 specific values. Each fan is constructed by using all 4,000 sets of the parameter and latent variable values from the MCMC output. Black (solid) line: the forecast yield curves with parameter certainty (PC) using the mean values of the MCMC output.

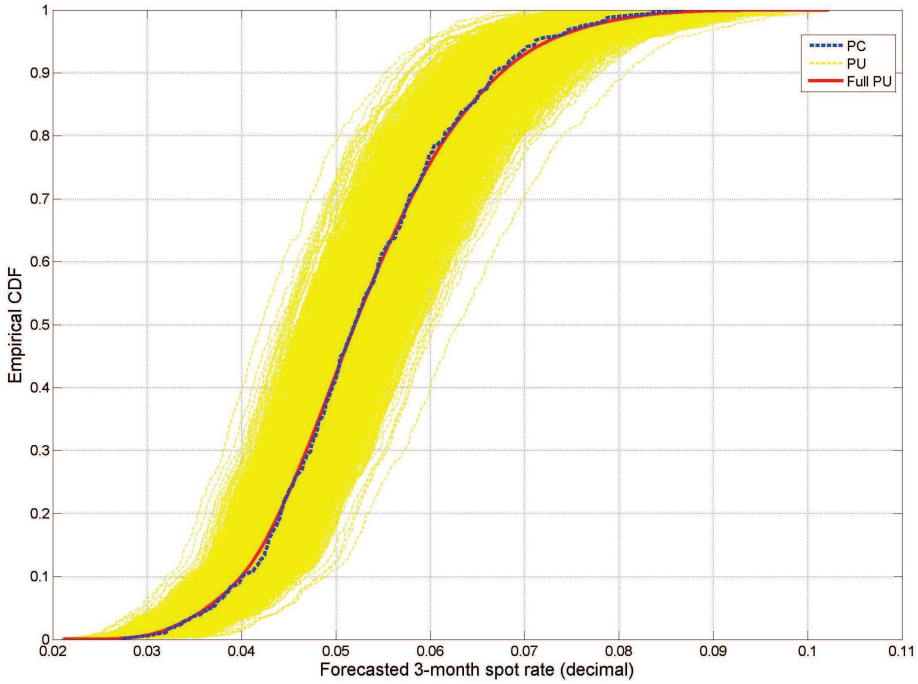

Figure 7.4: Distributions of the forecast 3-month spot rates for 1 year ahead with parameter uncertainty (PU) (yellow, dash) and parameter certainty (PC) (blue, dash). Each line uses one set of the parameter and latent variable values with "same" 500 values of future normal randomness Z_1 and Z_2. Red (solid) line: the empirical CDF for all 4,000 $\times$ 500 values of PU case.

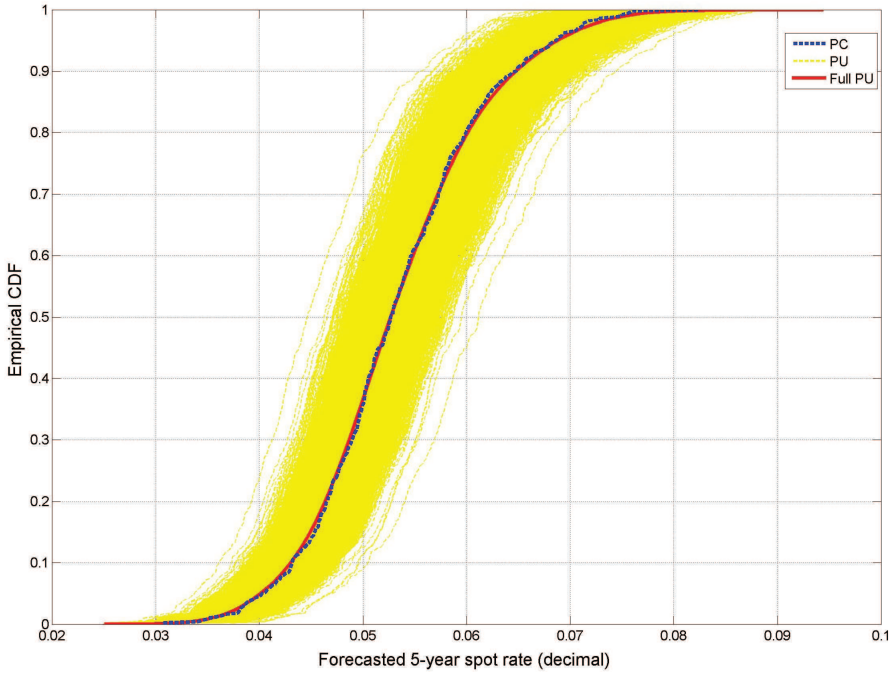

Figure 7.5: Distributions of the forecast 5-year spot rates for 1 year ahead with parameter uncertainty (PU) (yellow, dash) and parameter certainty (PC) (blue, dash). Each line uses one set of the parameter and latent variable values with "same 500" values of future normal randomness Z_1 and Z_2. Red (solid) line: the empirical CDF for all 4,000 $\times$ 500 values of PU case.

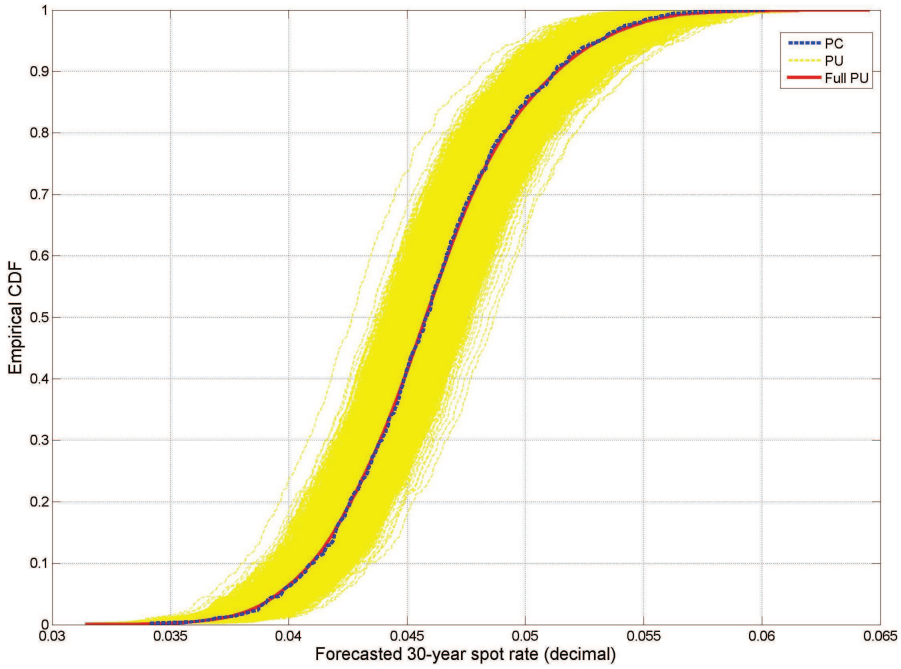

Figure 7.6: Distributions of the forecast 30-year spot rates for 1 year ahead with parameter uncertainty (PU) (yellow, dash) and parameter certainty (PC) (blue, dash). Each line uses one set of the parameter and latent variable values with "same" 500 values of future normal randomness Z_1 and Z_2. Red (solid) line: the empirical CDF for all 4,000 $\times$ 500 values of PU case.

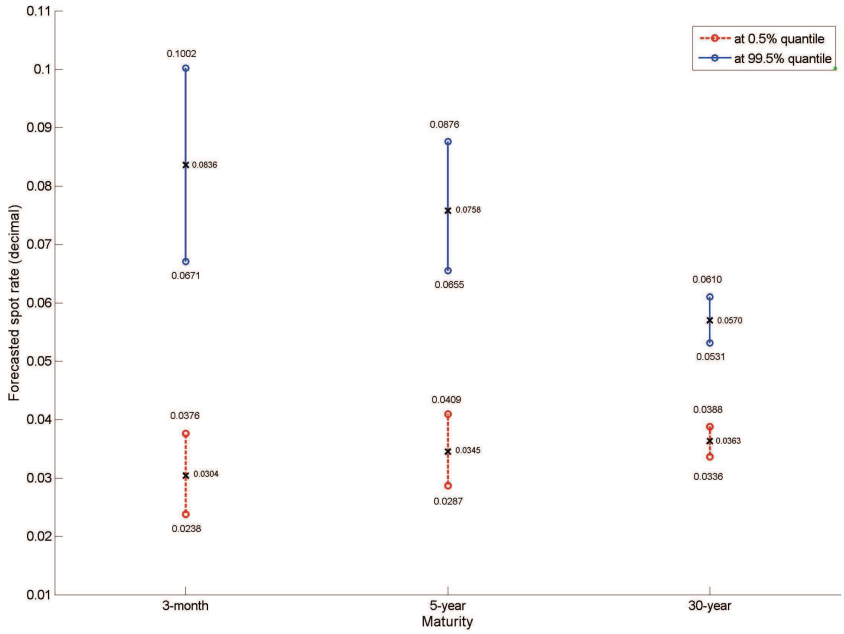

Figure 7.7: 0.5% and 99.5% quantile lines of the forecast 3-month, 5-year and 30-year spot rates for 1 year ahead with parameter uncertainty (PU). Cross mark: the values by parameter certainty (PC).

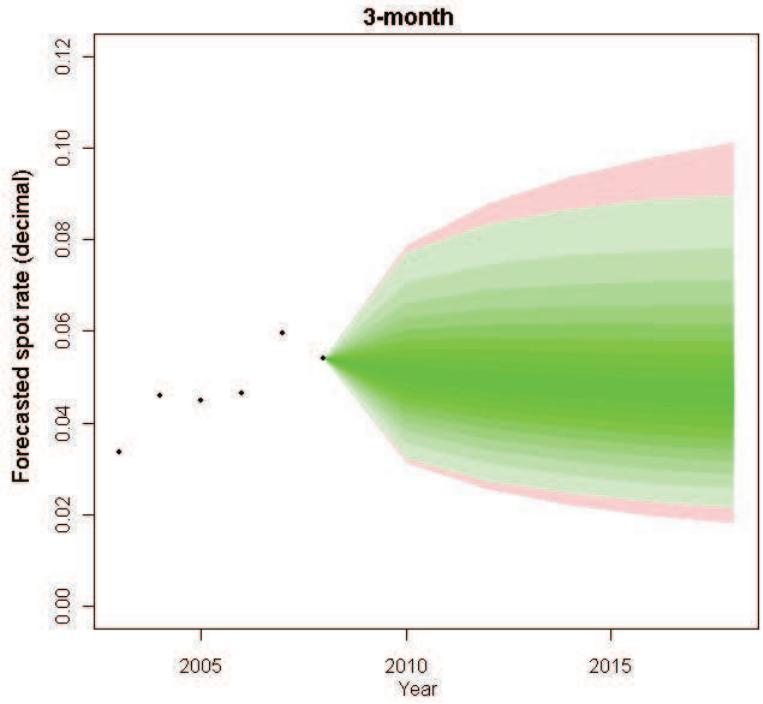

Figure 7.8: Fan charts for the forecast 3-month spot rates for 10 years ahead. Each set of the parameter and latent variable values is incorporated with future normal randomness Z_1 and Z_2 of 100 values for PU case (red fan) and 400,000 values for PC case (green fan). Dots: historical data.

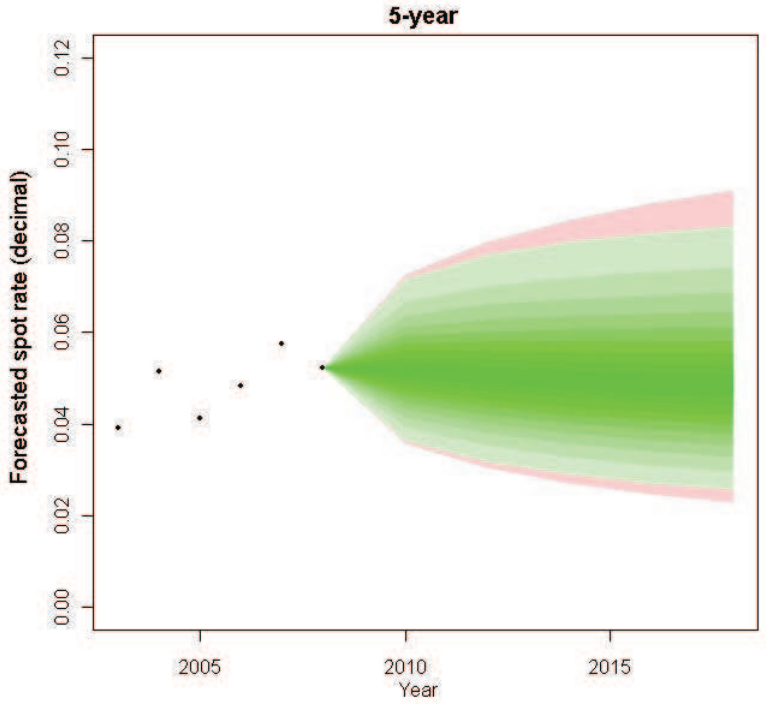

Figure 7.9: Fan charts for the forecast 5-year spot rates for 10 years ahead. Each set of the parameter and latent variable values is incorporated with future normal randomness Z_1 and Z_2 of 100 values for PU case (red fan) and 400,000 values for PC case (green fan). Dots: historical data.

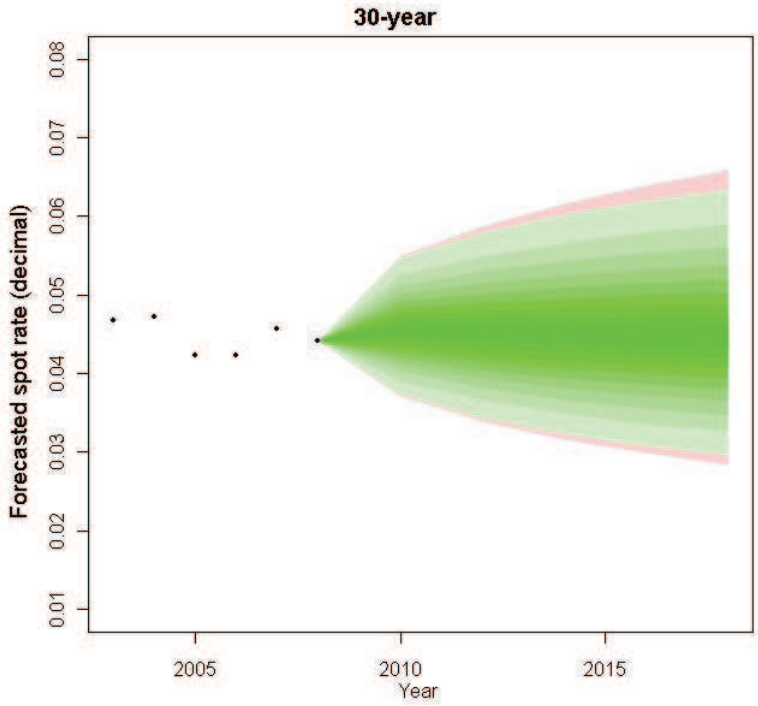

Figure 7.10: Fan charts for the forecast 30-year spot rates for 10 years ahead. Each set of the parameter and latent variable values is incorporated with future normal randomness Z_1 and Z_2 of 100 values for PU case (red fan) and 400,000 values for PC case (green fan). Dots: historical data.

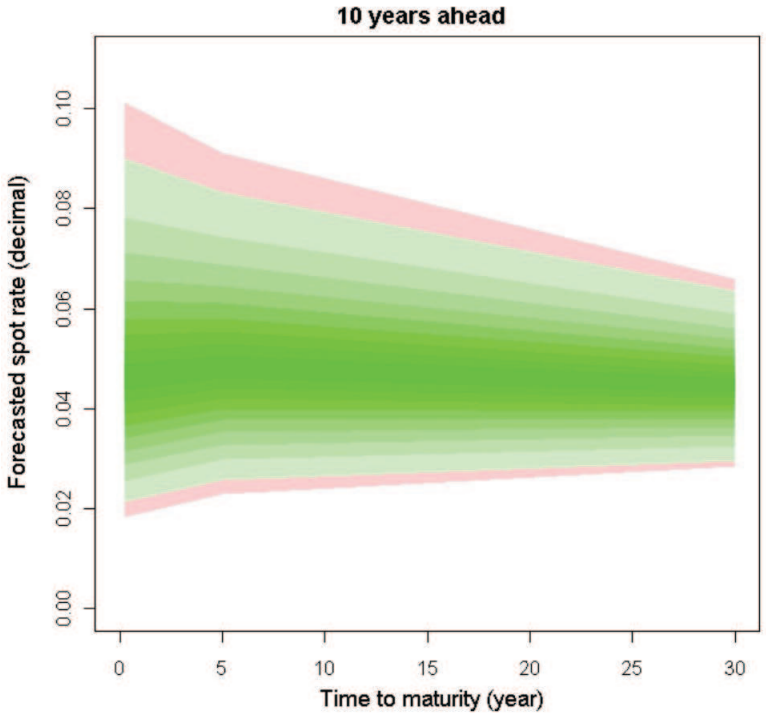

Figure 7.11: Fan charts for the forecast spot rates at 10 years ahead. Each set of the parameter and latent variable values is incorporated with future normal randomness Z_1 and Z_2 of 100 values for PU case (red fan) and 400,000 values for PC case (green fan).

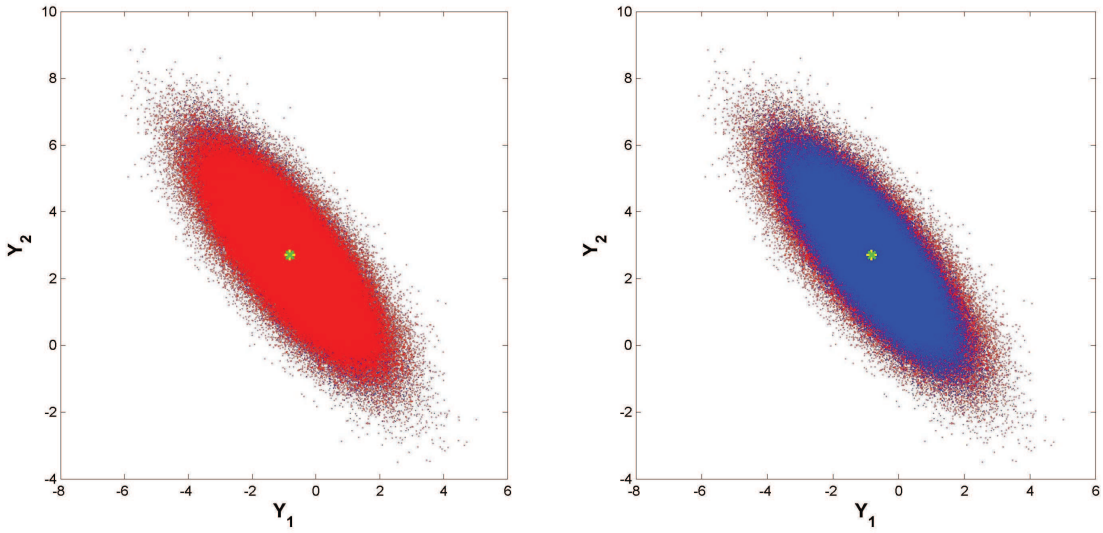

Figure 7.12: Scatter plots of the forecast latent variables $Y_1(t)$ and $Y_2(t)$ at 10 years ahead by parameter uncertainty (PU) and parameter certainty (PC). Left: PU case (red dot, 4,000 $\times$ 100 values) overlays PC case (blue dot, 1 $\times$ 400,000 values). Right: PC case overlays PU case. Plus mark: the mean for PC case. Cross mark: the mean for PU case.

Chapter 8

A Comparison of Fitting Two-Factor Cairns and Vasicek Term Structure Models

In this chapter, a two-factor Vasicek term structure model is first described and then estimated using the same framework, algorithm and UK Strips data as with the Cairns model in Chapter 6. Eventually, the goodness of fit, the forecast spot rates and annuity values from both models are investigated for comparison.

8.1 Two-Factor Vasicek Term Structure Model

The model discussed below is adapted from that proposed by Babbs and Nowman (1999). (Babbs and Nowman's model starts with the real world measure P and moves implicitly to the risk-neutral measure Q, whereas here we start with Q and move to P. This results in a simpler pricing formula.) The model is not the most general two-factor Vasicek model but the small number of restrictions result in a model that has similar elements in its structure to the two-factor Cairns model.

Suppose that $(\Omega, \mathcal{F}, P)$ is a probability space and $W(t)$ is a two-dimensional Wiener process adapted to a filtration $(\mathcal{F}_t)_{t\geq 0}$. For the two-factor Vasicek model, the short rate under the risk-neutral measure Q (equivalent to the real world measure P) can be defined by

$$r(t) = \mu + X_1(t) + X_2(t), \tag{8.1}$$

where $X_1(t)$ and $X_2(t)$ are the latent state variables follow

$$\begin{aligned} dX_1(t) &= -\alpha_1 X_1(t)dt + \sigma_1 d\tilde{W}_1(t) \\ dX_2(t) &= -\alpha_2 X_2(t)dt + \sigma_2 \rho d\tilde{W}_1(t) + \sigma_2\sqrt{1-\rho^2}d\tilde{W}_2(t), \end{aligned}$$

where $\tilde{W}_1(t)$ and $\tilde{W}_2(t)$ are two independent standard Wiener processes under the risk-neutral pricing measure Q and $\mu, \alpha_1, \alpha_2, \sigma_1, \sigma_2, \rho$ are constants.

Under this model, the price at t for £1 payable at $t+\tau$ is given by

$$V(\tau, X(t), \theta) = \exp[A(\tau) - \tau B(\alpha_1\tau)X_1(t) - \tau B(\alpha_2\tau)X_2(t)] \tag{8.2}$$

where $B(x) = (1-e^x)/x$ and

$$\begin{aligned} A(\tau) &= -\tau\mu + \frac{\tau\sigma_1^2}{2\alpha_1^2}(1 + B(2\alpha_1\tau) - 2B(\alpha_1\tau)) \\ &+ \frac{\tau\sigma_2^2}{2\alpha_2^2}(1 + B(2\alpha_2\tau) - 2B(\alpha_2\tau)) \\ &+ \frac{\tau\sigma_1\sigma_2\rho}{\alpha_1\alpha_2}(1 - B(\alpha_1\tau) - B(\alpha_2\tau) + B((\alpha_1+\alpha_2)\tau)). \end{aligned}$$

Note that the long-term spot rate $R(t, t+\tau)$ as $\tau \to \infty$ (which we shall denote by $R(t,\infty)$ and is equivalent to β in the Cairns model) is

$$R(t,\infty) = \mu - \frac{\sigma_1^2}{2\alpha_1^2} - \frac{\sigma_2^2}{2\alpha_2^2} - \frac{\sigma_1\sigma_2\rho}{\alpha_1\alpha_2}.$$

Dynamics under P are governed by

$$\begin{aligned} dX_1(t) &= -\alpha_1 X_1(t)dt + \sigma_1(dW_1(t) + \delta_1 dt) \\ &= -\alpha_1(X_1(t) - \frac{\sigma_1\delta_1}{\alpha_1})dt + \sigma_1 dW_1(t) \end{aligned}$$

and

$$dX_2(t) = -\alpha_2\left(X_2(t) - \frac{\sigma_2}{\alpha_2}(\delta_1\rho + \delta_2\sqrt{1-\rho^2})\right)dt + \sigma_1(\rho dW_1(t) + \sqrt{1-\rho^2}dW_2(t)),$$

where $W_1(t)$ and $W_2(t)$ are standard Wiener processes under the real world measure P and δ_1 and δ_2 are the corresponding market prices of risk.

In Chapter 5, γ_1 and γ_2 are the mean reversion levels for X_1 and X_2. Hence, we define

$$\begin{aligned}\gamma_1 &= \frac{\sigma_1\delta_1}{\alpha_1} \\ \gamma_2 &= \frac{\sigma_2}{\alpha_2}(\delta_1\rho + \delta_2\sqrt{1-\rho^2}).\end{aligned}$$

More generally, a key qualitative difference between the two-factor Vasicek model and the Cairns model is that the former model allows interest rates to become negative. Future spot rates under the Vasicek model are normally distributed whereas future spot rates under the Cairns model have a positively skewed distribution.

8.2 Estimation Results of Two-Factor Vasicek Model on Monthly UK Strips Data

Similar to the estimation framework for the Cairns model in Chapter 5, we suppose that interest rates in the market follow the two-factor Vasicek model in (8.2), i.e. the observations

$$P(t,\tau_{tj}) = V(\tau_{tj}; X(t), \theta) + \varepsilon(t,j), \tag{8.3}$$

where $\varepsilon(t,j) \sim$ i.i.d. $N(0,\sigma_\varepsilon^2)$. Accordingly, the full posterior distribution of the Vasicek bond price can be obtained in a similar form as (5.21) merely by replacing the function $C(\tau_{tj}; X(t), \theta)$ by $V(\tau_{tj}; X(t), \theta)$. In order to get the best comparison of estimation with the achieved results in Chapter 6, we also define the prior distributions for $\mu, \alpha_1, \alpha_2, \sigma_1, \sigma_2$ in the same way such that prior means are specified based on the posterior means from earlier simulations with about the same coefficient of variation.

In the following results, we run MCMC for 240,000 iterations by recording values for every 20th iteration using adaptive Metropolis-Hastings with a blocking strategy (exactly same algorithm as developed previously in Chapter 6). More precisely, we update groups of $(\alpha_1, \sigma_1, \rho), (\alpha_2, \sigma_2, \mu), (\gamma_1, \gamma_2)$ and $(X_1(t), X_2(t))$ for each time t where σ_ε is updated individually with a constant proposal variance. This is the same blocking as was used for the Cairns model and seems to be just as effective.

Figure 8.1 shows sample paths of all model parameters and the corresponding posterior estimates are provided in Table 8.1. As can be observed, all parameters generally converge very well, especially γ_1, γ_2, ρ and σ_ε. Comparing to the Cairns model, ρ turns to be much easier to estimate since Vasicek bond prices appear in a linear (affine) function of latent variables. The estimated ρ from the Vasicek model is slightly higher than Cairns model, but both are still strongly negative. Moreover, the long-term spot rate $R(t, \infty)$ (equivalent to β in the Cairns model) is also provided in Figure 8.2. It can be seen that the Vasicek model has perspective on the long-term rate higher than the Cairns model around 1% according to the estimated means.

Figure 8.3 demonstrates sample paths of latent variables for some selected months. As expected, all the chains converge reasonably well. Furthermore, when considering plots of 95% credible intervals for all t in Figure 8.4, we can find that they also behave similar to those from the Cairns model in Figure 6.6.

Finally, we consider the correlation structure for model parameters (Figures 8.5 and Table 8.2). For the Vasicek model, we can only notice strongly positive correlation for (α_1, σ_1), and moderately negative correlations for (σ_1, σ_2) and (γ_1, γ_2). Apart from these, strong correlations are not observed.

	Mean	Std.	95% Credible Interval	Acceptance rate	Scaling of the proposal std.
α_1	0.0386	0.00243	(0.03383, 0.04310)	11.31%	2.0
σ_1	0.0081	0.00024	(0.00768, 0.00858)	11.31%	2.0
ρ	-0.718	0.0449	(-0.7948, -0.6182)	11.31%	2.0
α_2	0.132	0.0037	(0.1258, 0.14039)	10.71%	1.8
σ_2	0.0136	0.00065	(0.01242, 0.01485)	10.71%	1.8
μ	0.0491	0.00045	(0.04826, 0.05006)	10.71%	1.8
γ_1	0.0109	0.02701	(-0.04223, 0.06335)	51.20%	1.0
γ_2	-0.0146	0.02176	(-0.05769, 0.02837)	51.20%	1.0
σ_ε	0.0024	0.00005	(0.00233, 0.00253)	44.35%	1.0
$R(t, \infty)$	0.0372	0.00182	(0.03323, 0.04035)	-	-

Table 8.1: Summary statistics of parameter posterior estimates of the two-factor Vasicek term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. The inference is made from 12,000 values (every 20th iteration out of 240,000 iterations).

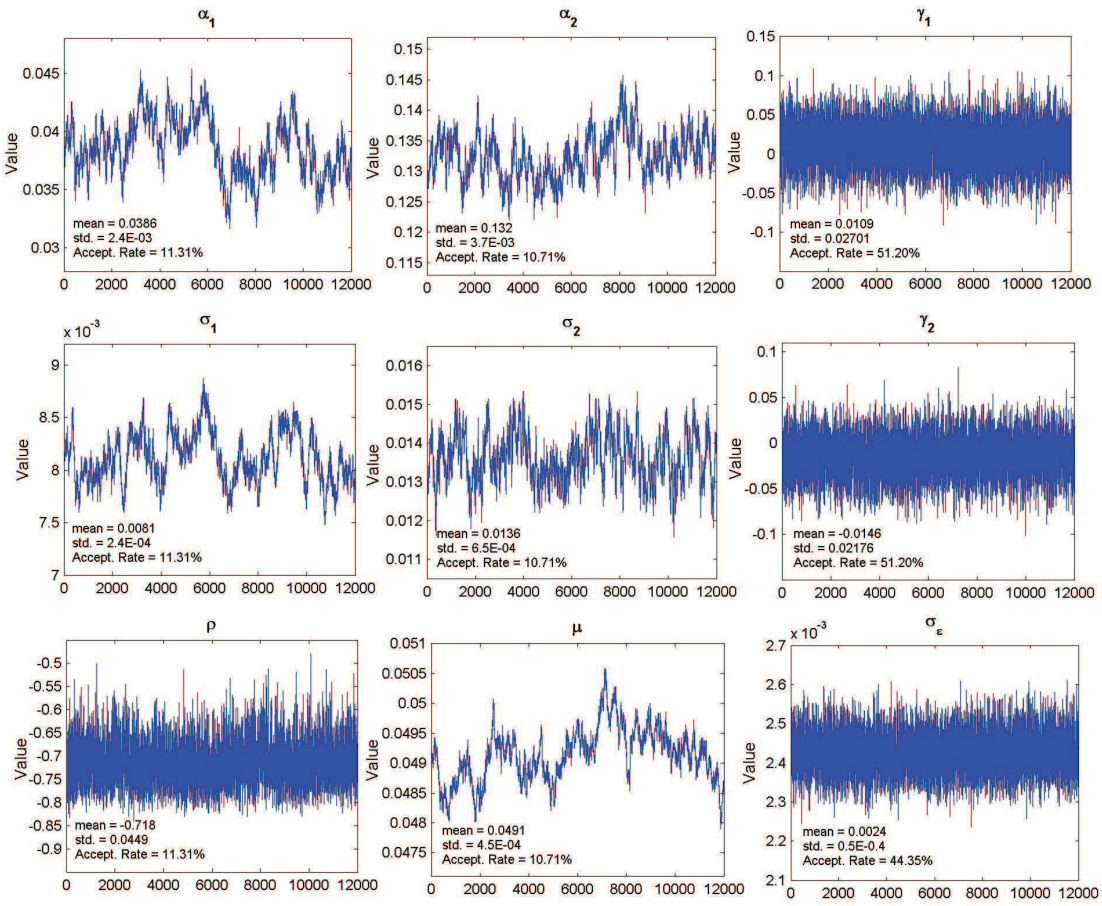

Figure 8.1: Sample paths of model parameters of the two-factor Vasicek term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. Plots are of 12,000 values (every 20th iteration out of 240,000 iterations).

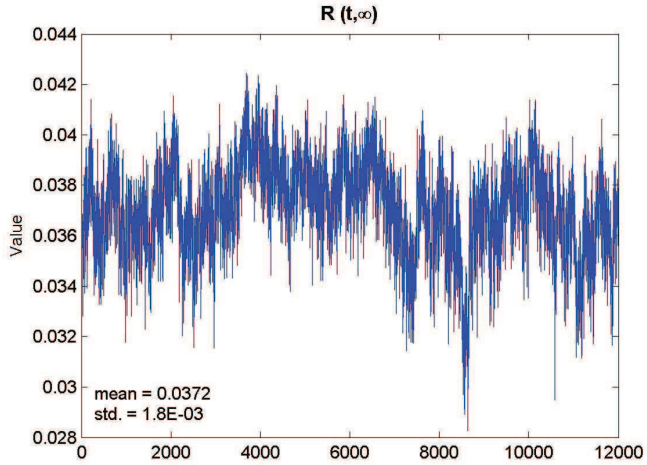

Figure 8.2: Sample paths of the long-term spot rate $R(t, \infty)$ (equivalent to β in the Cairns model) of the two-factor Vasicek term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. Plots are of 12,000 values (every 20th iteration out of 240,000 iterations).

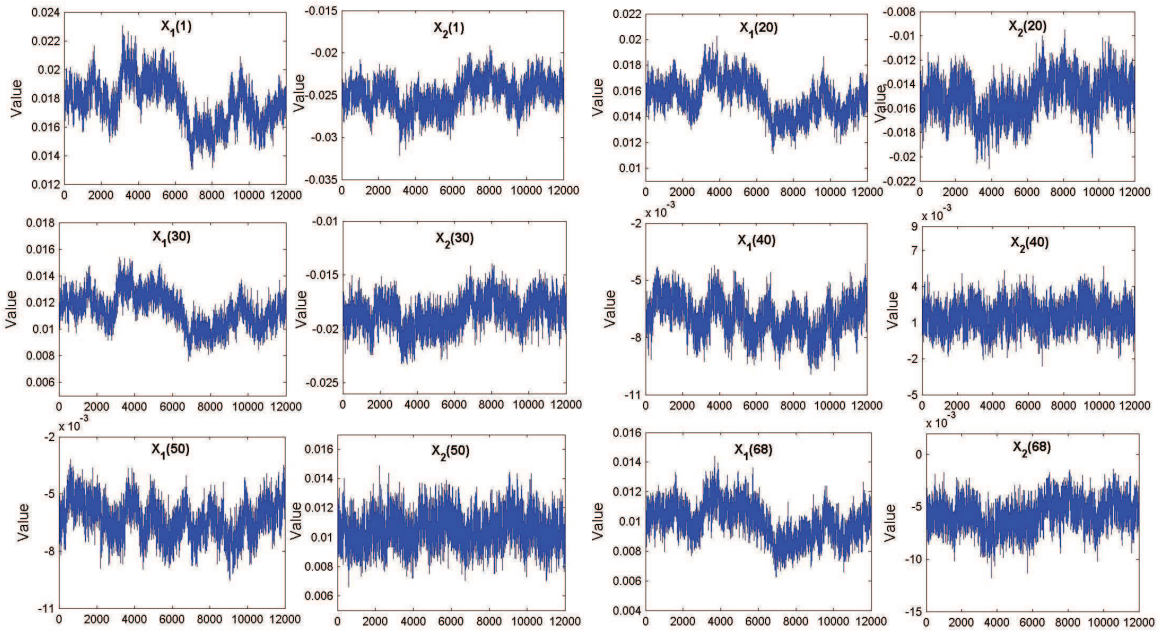

Figure 8.3: Sample paths of latent variables (for $t = 1, 20, 30, 40, 50$ and 68) of the two-factor Vasicek term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. Plots are of 12,000 values (every 20th iteration out of 240,000 iterations).

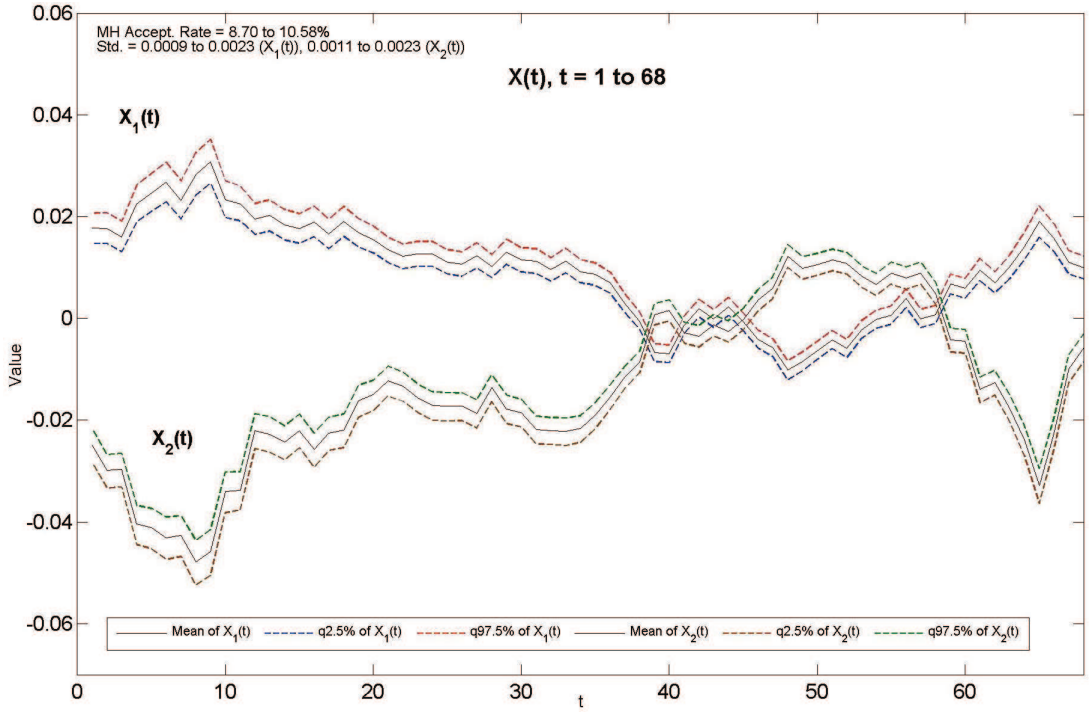

Figure 8.4: Plots of 95% credible interval constructed from the sample paths with the mean values of $X_1(t)$ and $X_2(t)$ for $t = 1, \ldots, 68$, of the two-factor Vasicek term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. The inference is made from 12,000 values (every 20th iteration out of 240,000 iterations).

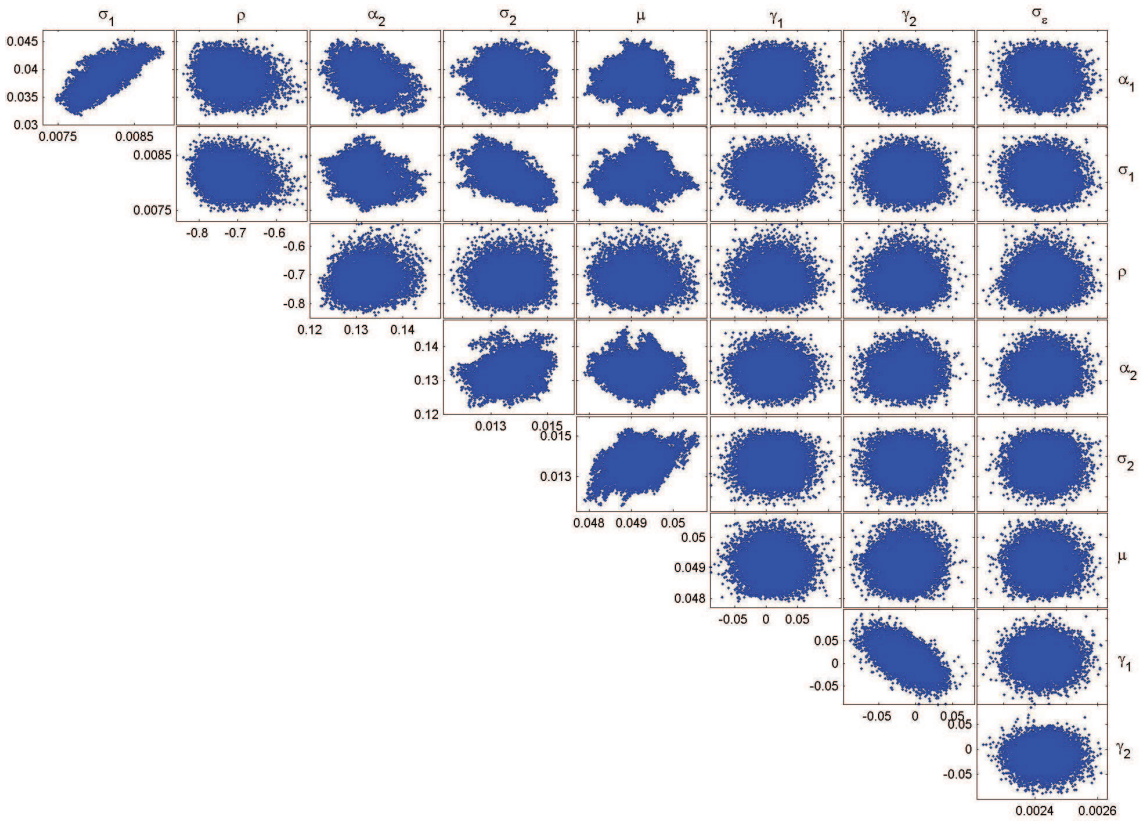

Figure 8.5: Scatter plots of model parameters of the two-factor Vasicek term structure model using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. Plots are of 12,000 values (every 20th iteration out of 240,000 iterations).

	α_1	σ_1	ρ	α_2	σ_2	μ	γ_1	γ_2	σ_ε	$X_1(t)$	$X_2(t)$
α_1	1.00	0.75	-0.15	-0.39	-0.13	-0.07	0.05	-0.05	-0.04	-0.45 to 0.86	-0.85 to 0.76
σ_1		1.00	-0.18	-0.26	-0.52	0.12	0.02	-0.03	-0.04	-0.61 to 0.52	-0.55 to 0.50
ρ			1.00	0.14	0.03	-0.07	-0.02	0.01	0.02	-0.14 to 0.13	-0.12 to 0.16
α_2				1.00	0.31	-0.03	-0.03	0.04	0.04	-0.65 to 0.32	-0.54 to 0.57
σ_2					1.00	0.33	0.01	0.00	0.02	-0.11 to 0.33	-0.45 to 0.01
μ						1.00	-0.03	0.03	0.00	-0.58 to -0.26	-0.37 to 0.09
γ_1							1.00	-0.59	0.00	0.01 to 0.07	-0.06 to 0.06
γ_2								1.00	0.00	-0.07 to 0.00	-0.07 to 0.07
σ_ε									1.00	-0.05 to 0.05	-0.05 to 0.06

Table 8.2: Correlation matrix of the simulation using the adaptive MH algorithm with the reparameterised posterior and re-evaluated priors, with monthly UK Strips data. The inference is made from 12,000 values (every 20th iteration out of 240,000 iterations).

8.3 Goodness of Fit: Two-Factor Cairns versus Vasicek Term Structure Models

In this section, we compare model fitting of the two-factor Cairns and Vasicek models using the estimated means of parameters and latent variables from the MCMC output. In terms of bond price and spot rate residuals (referring to equation 6.1), it can be seen from Table 8.3 that, overall, the Cairns model fits the data slightly better than the two-factor Vasicek model. For the sum of squared yield residuals by maturity, we found that those by Cairns model are lower than Vasicek model particularly for the first three maturities (0.25, 0.5, 1 years), which makes up for 62% of the total difference of average residual per maturity (0.224 out of 0.36 basis points).

	Total Sum of Squared Residuals (in decimal)		Average Residual per maturity (in bps)	
	Cairns	Vasicek	Cairns	Vasicek
Bond Price	0.007219	0.007227	23.04	23.05
Spot Rate	0.002282	0.002408	12.95	13.31

Table 8.3: Total sum of squared residuals and average residual from fitting the two-factor Cairns and Vasicek term structure model using the estimated means of parameters and latent variables form the MCMC output.

Figures 8.6a, 8.6b and 8.6c illustrate the fitted spot rate curves from the Cairns and Vasicek models for some selected months from November 2002 to June 2008 for comparison. Generally, distinct differences can be observed for the short-end, while the medium- and long-end of the spot rate curves from both models are very close.

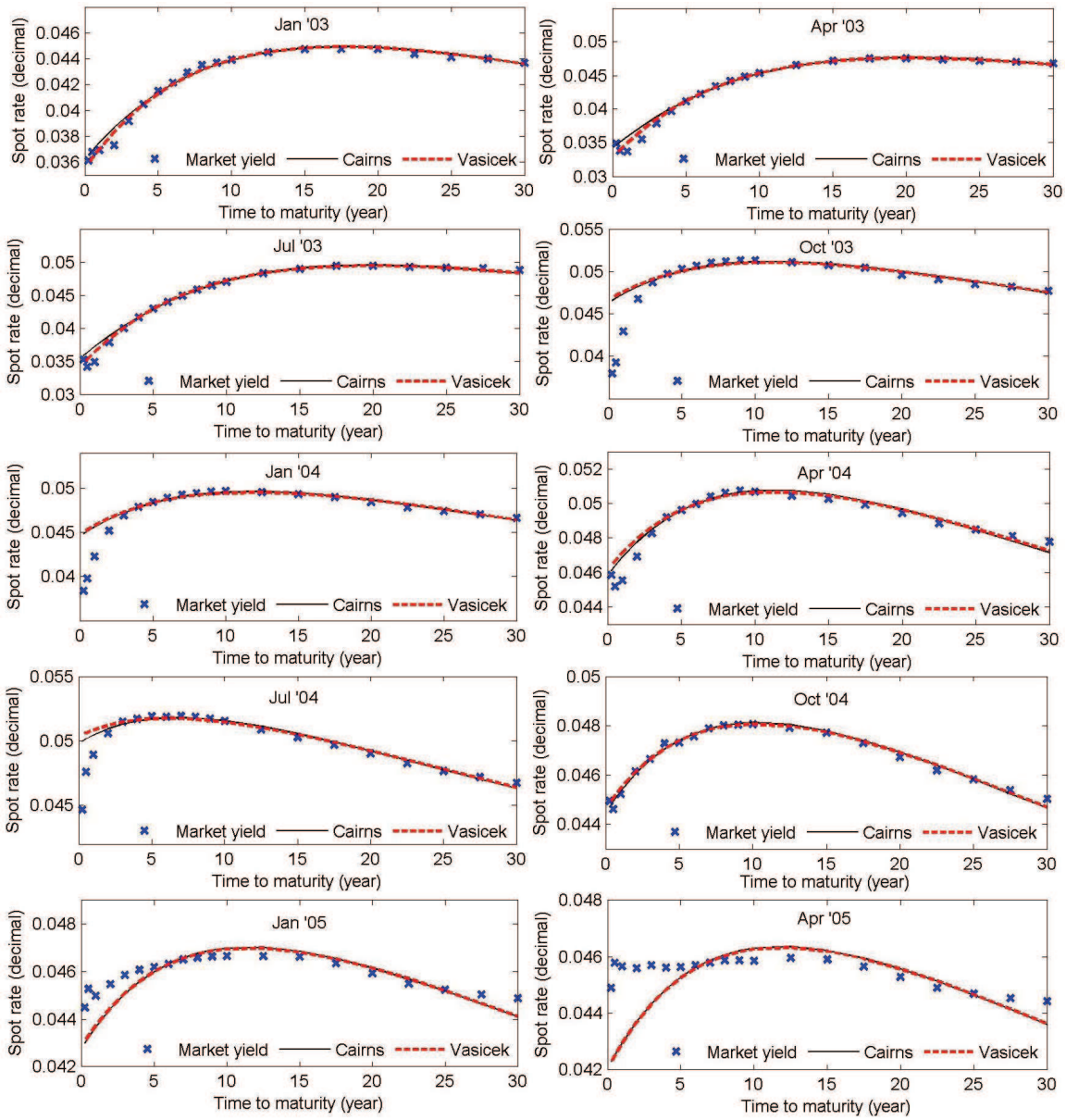

Figure 8.6a: The fitted spot rates of the two-factor Cairns model (black, solid) compared with the two-factor Vasicek model (red, dash) using the means of parameters and latent variables from the MCMC output for the selected months from November 2002 to April 2005. Cross marks: UK Strips yields.

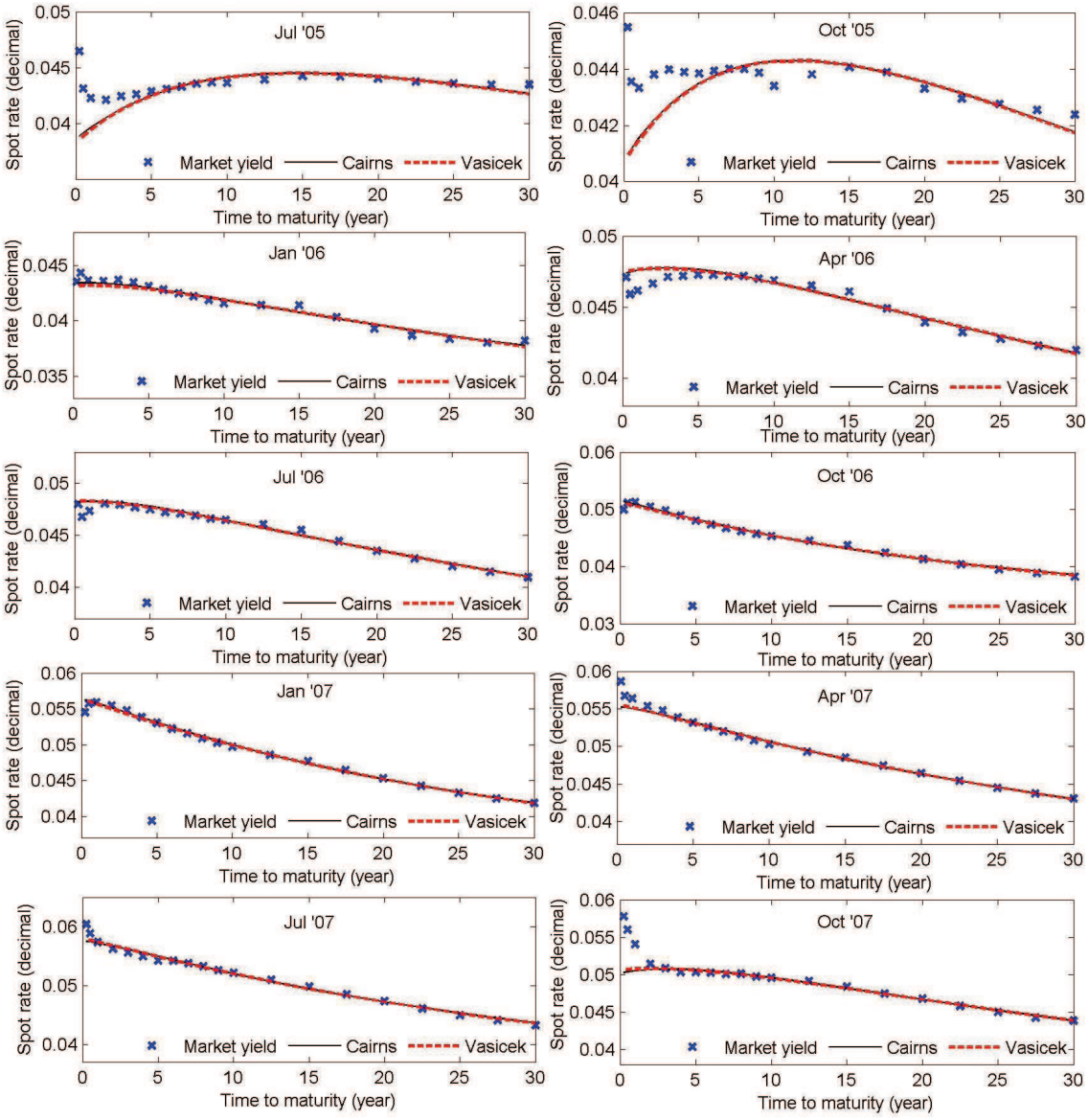

Figure 8.6b: The fitted spot rates of the two-factor Cairns model (black, solid) compared with the two-factor Vasicek model (red, dash) using the means of parameters and latent variables from the MCMC output for the selected months from May 2005 to October 2007. Cross marks: UK Strips yields.

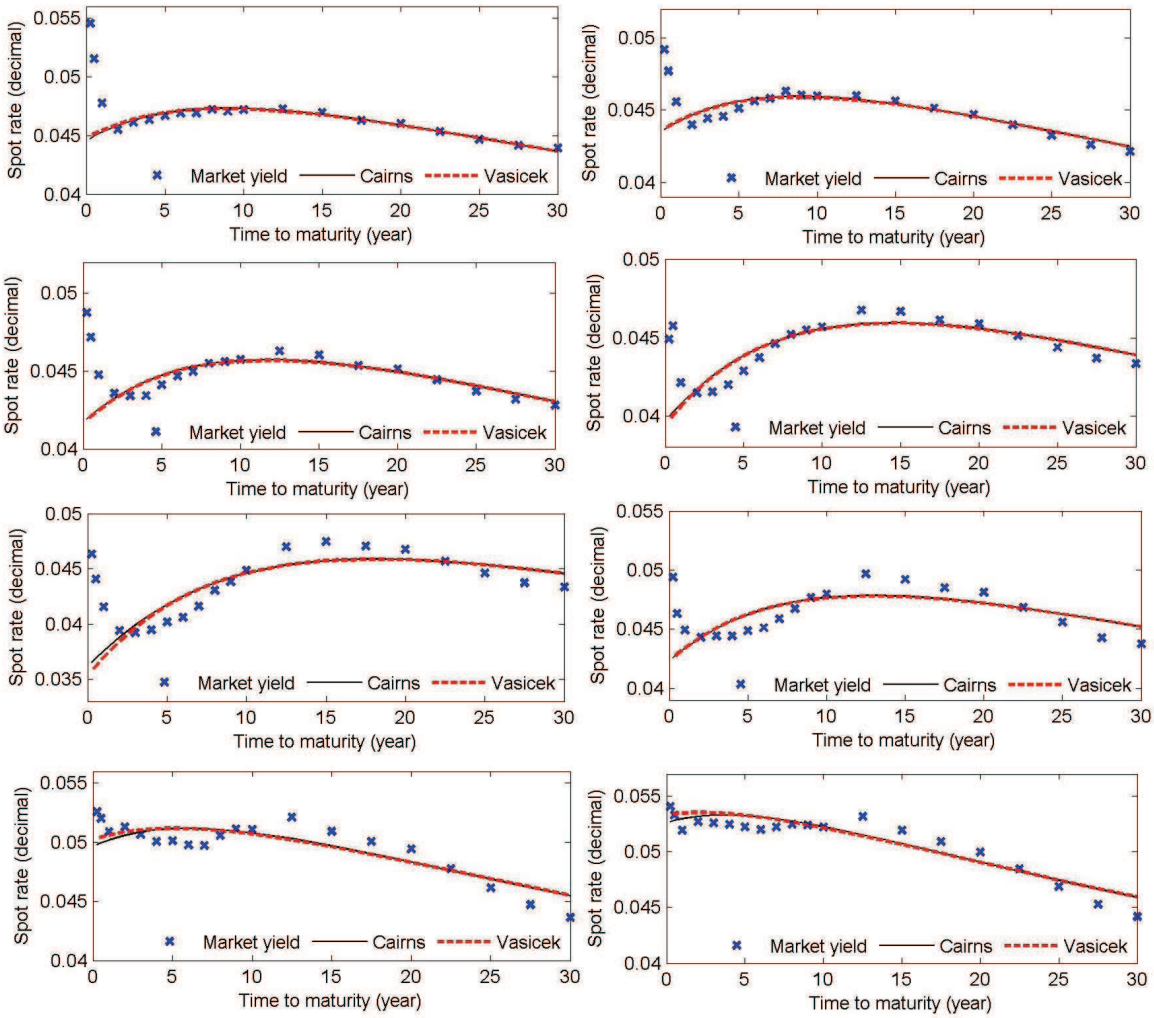

Figure 8.6c: The fitted spot rates of the two-factor Cairns model (black, solid) compared with the two-factor Vasicek model (red, dash) using the means of parameters and latent variables from the MCMC output for the last 8 months from November 2007 to June 2008. Cross marks: UK Strips yields.

8.4 Forecast Spot Rates and Annuity Prices: Two-Factor Cairns versus Vasicek Term Structure Models

In this section, we will compare forecast spot rates and annuity prices with parameter uncertainty (PU) and parameter certainty (PC) using achieved MCMC output from estimating the two-factor Cairns and Vasicek term structure models on monthly UK Strips data from November 2002 to June 2008. The simulation procedures are similar to those described in Chapter 7. For the PU case, 100 sets of parameter and latent variable values from MCMC output will be selected at random and incorporated with 100 fresh pairs of future normal randomness Z_1 and Z_2. For the PC case, the posterior means will be employed with 10,000 pairs of Z_1 and Z_2.

8.4.1 Forecasting Results: Spot Rates

Figures 8.7 and 8.8 compare forecast 3-month, 5-year and 30-year spot rates from the Cairns (blue line) and Vasicek (red line) models with PC (left column) and PU (right column) cases for 5 and 20 years ahead respectively. For 30-year maturity, the par yield curves (dash line) are also provided since in practice, long-dated par yields are referred to as much as spot rates. From the figures, observations can be made as follows.

- Despite achieving a very similar picture of model fitting from last section, forecast distributions from both models are noticeably different, especially for the short-term rates and longer time horizons. Most importantly, it can be observed that Vasicek model can generate negative rates for all three maturities which is an undesirable characteristic for term structure models.
- More consistent with historical data, the Cairns model can produce more realistic forecast spot rates in the sense that the higher rates can be obtained, albeit with a small probability.
- For 30-year maturity, the difference of forecast par yields between two models is clearer than spot rates, particularly on the right tails of distributions.
- Similar to the results in Chapter 7, parameter uncertainty does have an impact on forecast spot rates in all cases, particularly when time horizon is longer.

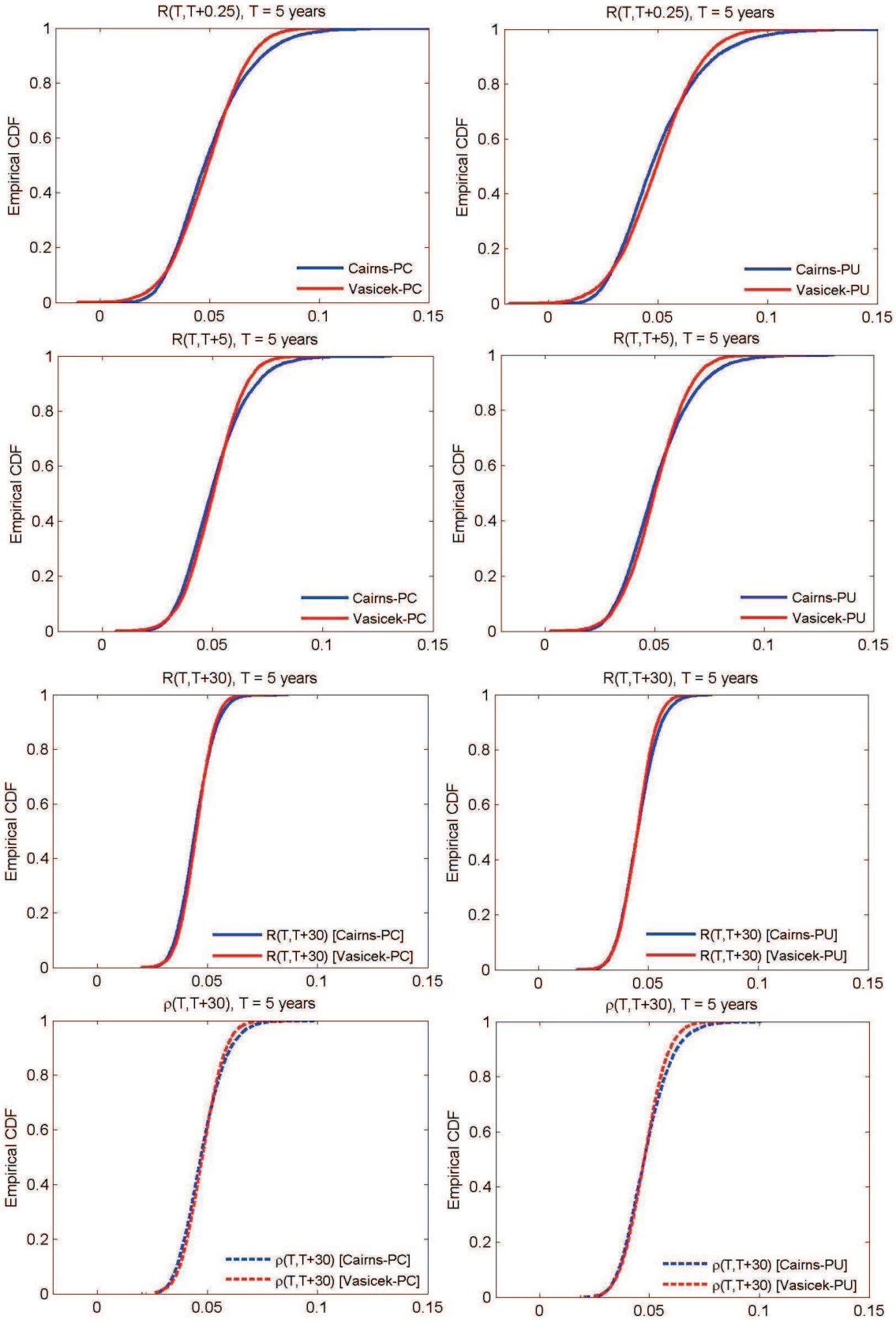

Figure 8.7: Distributions of the forecast 3-month, 5-year and 30-year spot rates (including 30-year par yields) for 5 years ahead with parameter certainty (PC) (left column) and parameter uncertainty (PU) (right column). Blue line: the two-factor Cairns model. Red line: the two-factor Vasicek model.

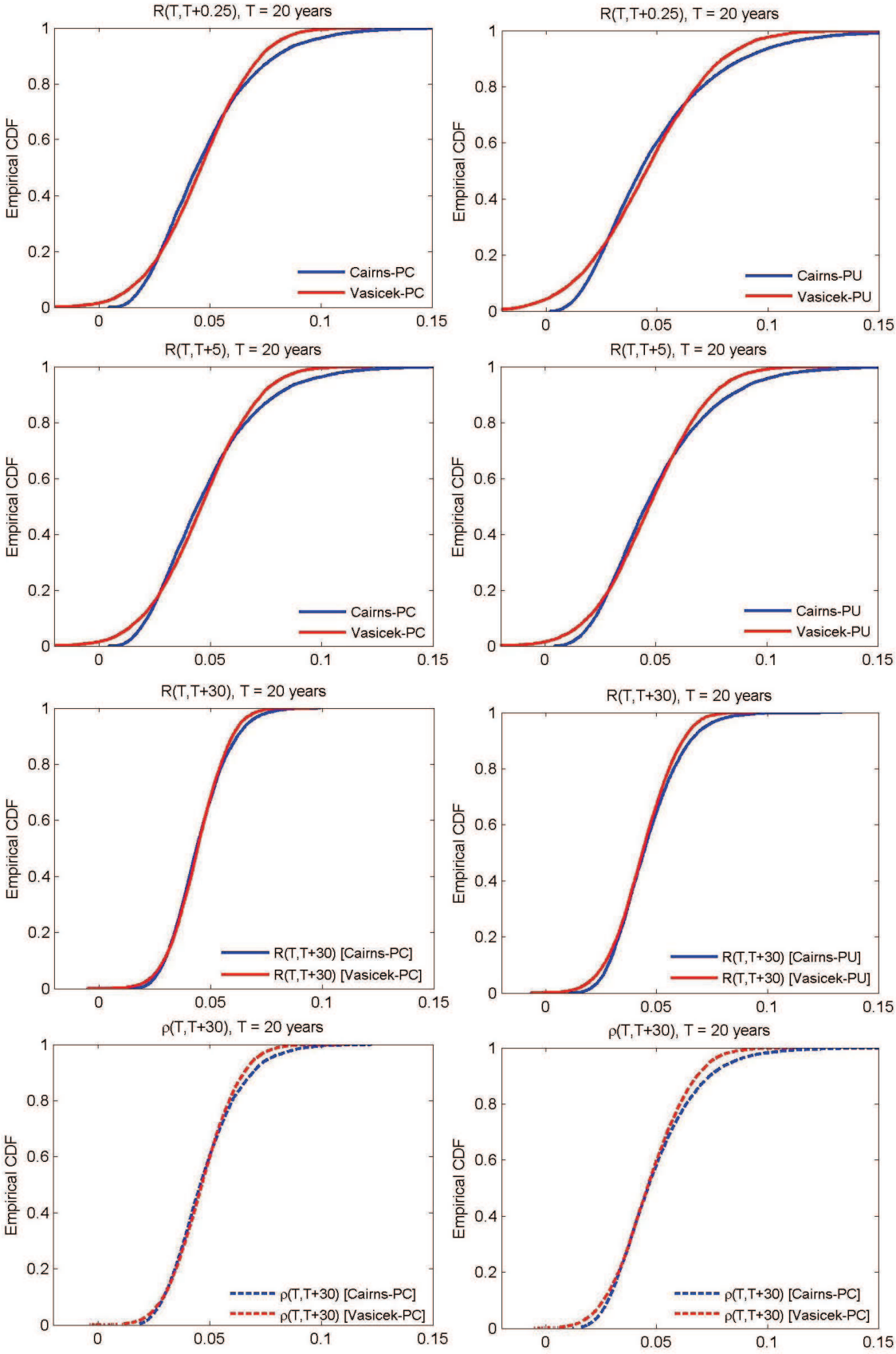

Figure 8.8: Distributions of the forecast 3-month, 5-year and 30-year spot rates (including 30-year par yields) for 20 years ahead with parameter certainty (PC) (left column) and parameter uncertainty (PU) (right column). Blue line: the two-factor Cairns model. Red line: the two-factor Vasicek model.

8.4.2 Forecasting Results: Annuity Prices

In order to have a clearer picture of the impact of the forecast rates on a more tangible financial contract, we particularly consider here forecast annuity values at age 65 for T years ahead which can be defined by

$$a_{65}(T) = \sum_{\tau=1}^{\infty} {}_{\tau}p_{65} P(T, \tau) \tag{8.4}$$

where ${}_{\tau}p_{65}$ is the probability of survival from age 65 to $65 + \tau$ (taken from the PMA92C20, the Faculty of Actuaries and the Institute of Actuaries, 2002, page 112) and $P(T, \tau)$ is a forecast zero-coupon bond price at time T maturing at $T + \tau$. Hence, the forecast bond prices (i.e. spot rates) from the previous section can be directly used to compute forecast annuity prices.

Figures 8.9a and 8.9b present empirical distributions and kernel densities of the forecast annuity values for 5, 10, 20 and 40 years ahead with PC and PU cases from Cairns (blue line) and Vasicek (red line) models in comparison. The corresponding summary statistics are also provided in Table 8.4. According to the results, we can make the following points:

- In both PC and PU cases, empirical cumulative distributions of the forecast annuity values from the two models are more different on both tails when time horizon is longer, representing higher model risk. The closeness of the distributions from the two models for 5 years ahead is as expected since annuity prices are calculated using a large number of long-maturity bond prices in which we can see from the previous section that model selection has least impact on the forecast long-term rates.

- According to the kernel densities, when time horizon is longer, the forecast prices from the Vasicek model tend to be skewed to the right (more low/negative interest rates), while those from the Cairns model are skewed to the left (high interest rates; the skewness is slightly more obvious in the PU case than in the PC case). The results are intuitive according to the corresponding forecast spot rate distributions which are skewed in opposite directions, and indicate more upside and downside risks from the Cairns and Vasicek models respectively. Additionally, it can be noticed that, in all cases, forecast values from Vasicek

model are more concentrated around their means (more peaked) than those from the Cairns model particularly for the shorter forecast time horizons and when comparing PC to PU case.

- From Table 8.4, it can be seen that, in all cases, means of the annuity values are relatively close given the magnitudes of standard deviations. Despite having rather similar means and standard deviations, the differences are certainly not negligible, with further differences revealed in the shapes of the distributions in Figure 8.8. In both PC and PU cases, the Cairns model gives rise to higher standard deviations than the Vasicek model for short forecast time horizon. However, when the time horizon is longer, the difference is diminishing and eventually the standard deviations under the Vasicek model turns out to be larger for 40 years ahead. For each model, it is clear that the PU case provides higher standard deviations than the PC case for all time horizons. Moreover, the difference is also higher provided longer time horizons (more obvious for Vasicek model).

In summary, this chapter presented comparisons between the Vasicek and Cairns models in terms of model fitting, forecasting and pricing with parameter uncertainty and certainty using the same methodology and market data as described in Chapter 6 and 7. Evidently, the developed MCMC algorithm was also found to be very efficient to estimate the two-factor Vasicek term structure model. Using estimated posterior means, we noticed that the Cairns model provided slightly better fit to the market data than the Vasicek model in terms of sum of squared residuals, particularly for the short-end of the spot rate curves. For forecasting and pricing using the two models, distinct differences can be seen on the tails of the distributions in which the Vasicek model revealed a substantial pitfall of generating negative rates. Regarding forecast annuity prices, the Vasicek and Cairns model tended to have downside and upside risks respectively, indicating model risk for forecasting and pricing with particular consequences for the pricing, for example, of guaranteed annuity options. Parameter uncertainty was found to have a similar impact on forecasts made by the two models.

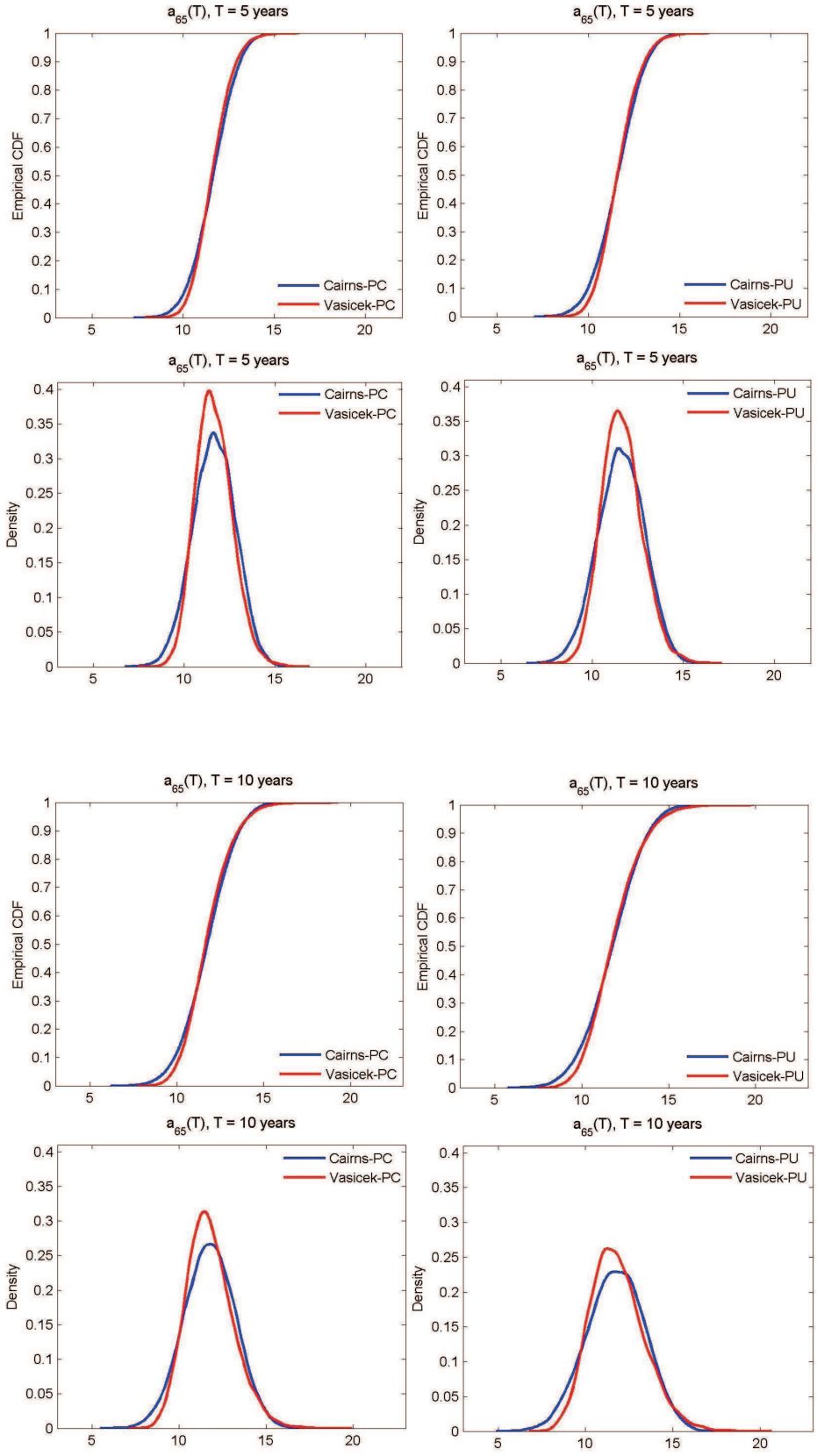

Figure 8.9a: Distributions and kernel densities of the forecast annuity values for 5 and 10 years ahead with parameter certainty (PC) (left column) and parameter uncertainty (PU) (right column). Blue (solid) line: the two-factor Cairns model. Red (solid) line: the two-factor Vasicek model.

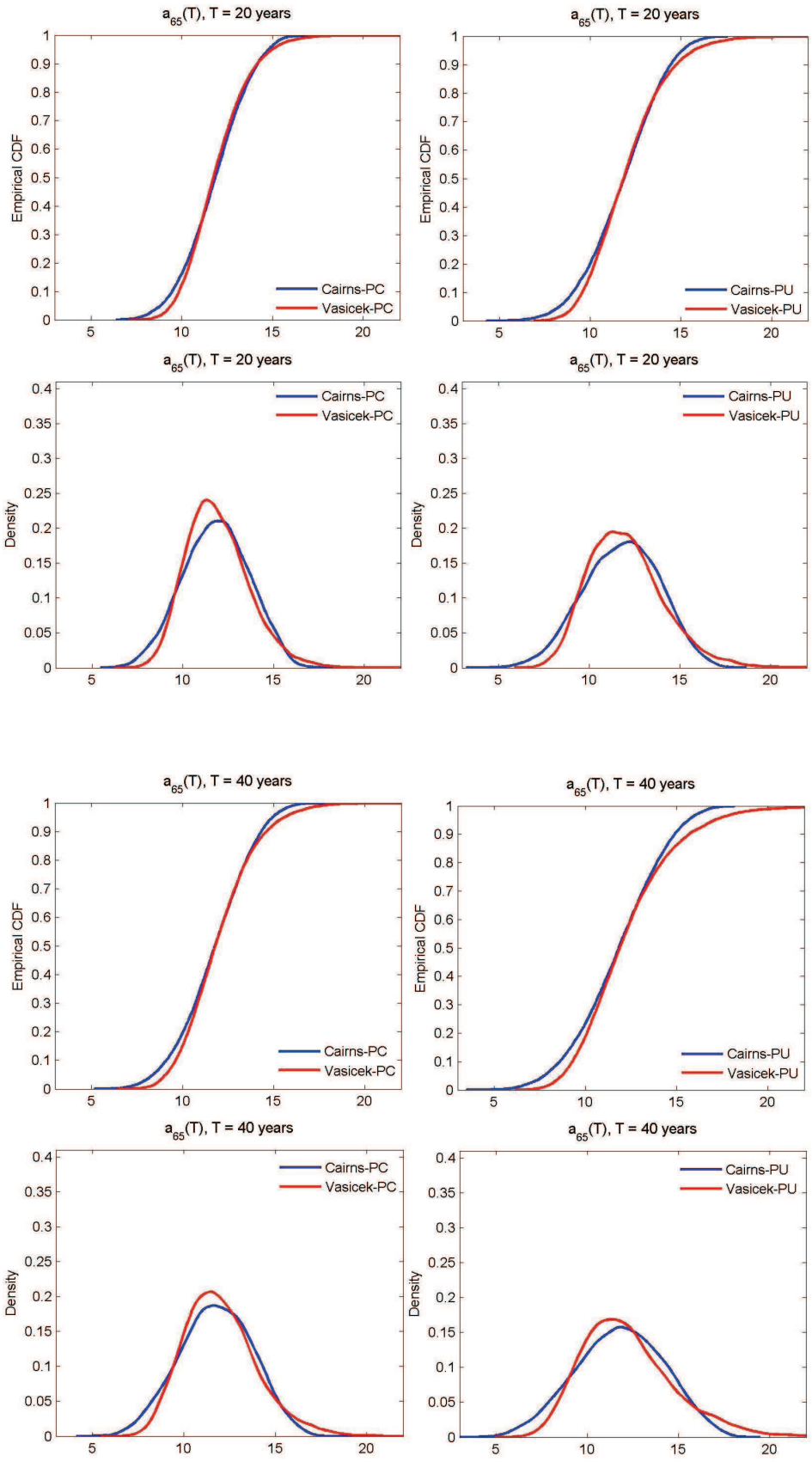

Figure 8.9b: Distributions and kernel densities of the forecast annuity values for 20 and 40 years ahead with parameter certainty (PC) (left column) and parameter uncertainty (PU) (right column). Blue (solid) line: the two-factor Cairns model. Red (solid) line: the two-factor Vasicek model.

Forecast Annuity Means and Standard Deviations

Forecast Time Horizon	Cairns Model		Vasicek Model	
	PC	PU	PC	PU
5 years ahead	11.641 (1.1586)	11.598 (1.2530)	11.632 (1.0078)	11.647 (1.0929)
10 years ahead	11.741 (1.4559)	11.725 (1.6460)	11.769 (1.3391)	11.832 (1.5459)
20 years ahead	11.827 (1.7935)	11.800 (2.0683)	11.915 (1.7181)	12.036 (2.0673)
40 years ahead	11.752 (1.9842)	11.794 (2.3933)	11.995 (1.9947)	12.261 (2.6289)

Table 8.4: Forecast annuity means and standard deviations for 5, 10, 20 and 40 years ahead by the two-factor Cairns and Vasicek term structure models with parameter certainty (PC) and parameter uncertainty (PU).

Chapter 9

Conclusions and Further Research

In this thesis, we have developed MCMC algorithms in order to estimate the two-factor Cairns term structure model using both simulated and market data. The main contribution of this thesis is therefore the development and use of MCMC simulation for estimating the latent state variables and model parameters of the Cairns model. The existence of latent variables is a common issue that causes difficulty to the estimation of many continuous-time term structure models, and here it can be effectively dealt with by using MCMC methodology under a Bayesian approach.

Methodology

There are three main methodological considerations that have been addressed and are crucial for the success of our MCMC implementation.

- *Exact solution of a two-dimensional Ornstein-Uhlenbeck process.* Although it is known that a simple Euler approximation of stochastic differential equations generally gives rise to poor estimation (especially when the step size is large), having borne this in mind we first still use the Euler method to discretise continuous-time processes of the latent variables for simplicity. As a consequence, it is the case that some parameter estimates are strongly biased. However, once the exact solution was derived and employed for the bond price posterior distribution, such biases were eliminated and the estimating results were significantly improved.

- *Reparameterised bond posterior and adaptive Metropolis-Hastings algorithm with a blocking strategy.* Due to the complexity of the model and evidence

of correlations among some particular parameters, the chains converge very slowly using the standard MH algorithm. Thus, we considered reparameterisation of some parameters in the bond posterior associated with using an adaptive Metropolis-Hastings algorithm with a blocking scheme (where candidate points are sampled from a multivariate normal proposal distribution). In turn, the convergence of model parameters was obviously improved. Even though it can be noticed that the correlation structure of some parameters has changed, it does not affect the estimating results and we can still obtain the posterior estimates of all model parameters and latent variables that are close to the true values when using the simulated data. Furthermore, the adaptive scheme also allows us to shorten the time of searching for suitable values for the constant proposal variances. More precisely, we are only required to determine the suitable numbers of sample paths for computing the proposal covariance matrix and its scaling factors which is found to be much easier than calibrating the constant proposal variances.

- *Numerical bond price routine in C++.* All the numerical integration methods used (Appendix C) result in very accurately approximated Cairns bond prices, and therefore the method selection should not provide any significant difference to our estimating results for the latent variables and model parameters. Nevertheless, for efficient implementation of MCMC simulation (which naturally takes a long time to run for a complex model), this part of programming turns to be the bottleneck which is extremely time consuming. Moving this routine of the algorithm from Matlab to C++ was a crucial step that made calibration of the Cairns model feasible. Without this, any progress would be hard to achieve since we may need to take much longer time to check results and then calibrate our algorithms accordingly.

Key Findings from the Main Results

According to the estimation and forecasting results using the simulated and UK Strips data in Chapter 5, 6, 7 and 8, the key findings can be summarised as follows.

- With our improved MCMC algorithm (the adaptive Metropolis-Hastings algorithm with reparameterised bond posterior and a blocking scheme), the chains

of model parameters and latent variables converge reasonably well for both datasets. However, we found that the uncertainty of MCMC estimates (e.g. coefficients of variation of the posterior estimates) using market data is much higher than using the simulated data. One possible reason is that the simulated data are less complex than the UK Strips which clearly have large deviation at the short-end, particularly for the last 8 months over the time period considered.

- By using the mean values of the parameters and latent variables from the MCMC output, the two-factor Cairns model is generally fitted to the UK Strips data fairly well for the medium- and long-term yields (except during the turmoil period after 2008), but is rather poorly fitted for the short-end of the yield curves. Despite knowing that discrepancies between the estimated and market yields is a common drawback of time-homogeneous arbitrage-free models, two factors may be sufficient in the Cairns model to capture the dynamics of the data in some cases. It is obvious though that at least one additional factor may be required in order to be able to capture the dynamics of UK Strips at the short-end in particular.

- Our assumption regarding the i.i.d.-normal residuals in the framework setting may not always be valid. It is evident that generally the residuals of the Cairns bond prices after fitting the model using UK Strips are correlated and some may also not normally distributed. Other error structures could also be considered in future research.

- Parameter uncertainty clearly does have an impact on the forecasting yield curves, particularly for the short-end. It becomes essentially important when we consider the distributions of the forecast interest rates at both tails (e.g. Value-at-Risk). Of all model parameters, the market prices of risk are likely to be influential parameters for the forecasting. Eventually, we may conclude that allowance for parameter uncertainty should not be neglected when using any model.

- Our developed MCMC algorithm was also found to be very efficient to estimate the two-factor Vasicek term structure model. The Cairns model fits the short-end of yield curves better than the Vasicek model in terms of sum of

squared residuals when using estimated posterior mean values of parameter and latent variables. Comparing their forecast spot rates, we can observe distinct differences on tails of the distributions where the Vasicek model discloses a substantial drawback of producing negative rates. Furthermore, model risk also reveals when considering the distributions of forecast annuity prices from the two models (negative skewness for the Cairns model and positive skewness for the Vasicek model).

Further Research

There is enough scope for further research. Firstly, the two-factor Cairns term structure model may be extended to the three-factor model in order to improve the model fitting at the short-end of the yield curves. Secondly, we may also consider the the Cairns model in a more general form with stochastic volatility (Cairns and Garcia Rosas, 2004). It is also interesting to calibrate the model using other real market data such as US data and then compare posterior parameter estimates and forecasting results. Also, some additional parameterisation for the bond posterior distribution may be considered for improving the convergence of the correlation parameter ρ. For the framework setting, the bond price residuals can be allowed to be correlated. Once the model is fitted reasonably well to the data, we may also attempt to use the model for derivative pricing.

Finally, our MCMC methodology is not limited only to the term structure models for interest rates. It may be adapted to any other models that have some similarity in terms of incorporating an Ornstein-Uhlenbeck process as the driving latent state variables.

References

[1] Ahn, A.-H., Dittmar, R.F. and Gallant, A.R. (2002). Quadratic term structure models: theory and evidence. *Review of Financial Studies*, **15**, 243-288.

[2] Alexander, C. (2002). Principal component models for generating large covariance matrices, *Review of Banking, Finance and Monetary Economics, Economic Notes*, **31:2**, 337-359.

[3] Andersen, T.G. and Lund, J. (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics*, **77**, 343-377.

[4] Anderson, N. and Sleath, J. (2001). New estimates of the UK real and nominal yield curves. *Bank of England's Working Paper Series*, **ISSN 1368-5562**.

[5] Babbs, S.H. and Nowman, K.B. (1999). Kalman Filtering of Generailized Vasicek Term Structure Models, *Journal of Financial and Quantitative Analysis*, **1**, 115-130.

[6] Ball, C.A. and Torous, W.N. (1996). Unit roots and the estimation of interest rate dynamics. *Journal of Empirical Finance*, **3**, 215-238.

[7] Bank for International Settlements (2005). Zero-coupon yield curves: technical documentation. *BIS Papers*, **25**.

[8] Bester, A.C. (2004). Random Fields and Affine Models for Interest Rates: An Empirical Comparison. *Working Paper*, Duke University.

[9] Black, F. and Karansinski, P. (1991). Bond and option pricing when short rates are log-normal. *Financial Analysts Journal*, **47** (July-August), 52-59.

[10] Bliss, R.R. (1997). Movements in the Term Structure of Interest Rates, *Federal Reserve Bank of Atlanta Economic Review*, Forth Quarter.

[11] Brace, A., Gatarek, D. and Musiela, M. (1997). The market model of interest rate dynamics. *Mathematical Finance*, **7**, 127-155.

[12] Brzeźniak, Z. and Zastawniak, T. (1998). *Basic Stochastic Processes*, Springer-Verlag: London.

[13] Cairns, A.J.G. (1998). Descriptive bond-yield and forward-rate models for the British government securities' market. *British Actuarial Journal*, **4**, 265-321.

[14] Cairns, A.J.G. (2000). A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics*, **27**, 313-330.

[15] Cairns, A.J.G. (2004a). A Family of term-structure models for long-term risk management and derivative pricing. *Mathematical Finance*, **14**, 415-444.

[16] Cairns, A.J.G. (2004b). *Interest-Rate Models: An Introduction*, NJ: Princeton University Press.

[17] Cairns, A.J.G. and Garcia Rosas, S.A. (2004). A Family Of Term-structure Models with Stochastic Volatility. *Presented at the 3rd World Congress of the Bachelier Society, Chicago, July 2004.*

[18] Chan, K.C., Karolyi, G.A., Longstaff, F.A. and Sanders, A.B. (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance*, **47**, 1209-1227.

[19] Cowles, M.K. and Carlin, B.P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *Journal of American Statistical Association*, **91**, 883-904.

[20] Cox, J., Ingersoll, J. and Ross, S. (1985). A theory of the term-structure of interest rates. *Econometrica*, **53**, 385-408.

[21] Dai, Q. and Singleton, K. (2000). Specification analysis of affine term structure models. *Journal of Finance*, **55**, 1943-1978.

[22] Davis, J. and Rabinowitz, P. (1967). *Numerical Integration*, Blaisdell Publishing Company.

[23] Diebold, F.X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, **130**, 337-364.

[24] Diebold, F.X., Rudebusch G.D. and Aruoba, S.B. (2006). The macroeconomy and the yield curve: a dynamic latent factor approach, *Journal of Econometrics*, **131**, 309-338.

[25] Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, **6**, 379-406.

[26] Duffie, D. and Singleton, K.J. (1993). Simulated moments estimation of Markov models of asset prices. *Econometrica*, **61**, 929-952.

[27] Duffee, G.R. (2002). Term premia and interest rate forecasts in affine models. *Journal of Finance*, **57**, 405-443.

[28] Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, **50**, 987-1008.

[29] Eraker, B. (2001). MCMC Analysis of Diffusion Models with Application to Finance. *Journal of Business & Economic Statistics*, **19**, 177-191.

[30] Fleasaker, B. and Hughston, L. (1996). Positive interest. *Risk*, **9**(1), 46-49.

[31] Gallant, A.R. and Tauchen, G. (1989). Seminonparametric estimation of conditionally constrained heterogeneous processes: asset pricing applications. *Econometrica*, **57**, 1091-1120.

[32] Gallant, A.R. and Tauchen, G. (1996). Which moments to match? *Econometric Theory*, **12**, 657-681.

[33] Gallant, A.R. and Tauchen, G. (2001). Efficient method of moments. Working Paper, University of North Carolina-Chapel Hill.

[34] Garmerman, D. and Lopes, H.F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference (2nd edn)*, Chapman & Hall.

[35] Gander, W. and Gautschi, W. (1998). Adaptive quadrature - revisited. *BIT*, **40**, 84-100.

[36] Haario, H., Laine M., Mira, A. and Saksman E. (2005). DRAM - efficient adaptive MCMC. *Statistics and Computing*, **16**, 339-354.

[37] Hansen, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**, 1029-1055.

[38] Heath, D., Jarrow, R. and Morton, A. (1992). Bond pricing and the term structure of interests: a new methodology for contingent claims valuation. *Econometrica*, **60**, 77-105.

[39] Heath, D. and Schweizer, M. (2000). Martingales versus PDEs in finance: an equivalence result with examples. *Journal of Applied Probability*, **37**, 947-957.

[40] Ho, S.Y. and Lee, S.-B. (1986). Term structure movements and pricing interest rate contingent claims. *Journal of Finance*, **41**, 1011-1029.

[41] Hu, H. (2005). Markov Chain Monte Carlo Estimation of Multi-Factor Affine Term-Structure Models. *Ph.D. Dissertation*, University of California Los Angeles.

[42] Hull, J.C. and White, A.D. (1990). Pricing interest rate derivative securities. *Review of Financial Studies*, **3**, 573-592.

[43] Imbens, G.W. (2002). Generalized method of moments and empirical likelihood. *Journal of Business & Economic Statistics*, **20**, 493-506.

[44] Jagannathan, R. and Wang, Z. (2002). Generalized method of moments: applications in finance. *Journal of Business & Economic Statistics*, **20**, 470-481.

[45] James, J. and Webber, N. (2000). *Interest Rate Modelling*, Chichester: Wiley.

[46] Lamoureux, C.G. and Witte, H.D. (2002). Empirical Analysis of the Yield Curve: The Information in the Data Viewed through the Window of Cox, Ingersoll, and Ross. *Journal of Finance*, **57**, 1479-1520.

[47] Litterman, R. and Scheinkman, J.A. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, **1**, 54-61.

[48] Longstaff, F.A. and Schwartz, E.S. (1992). Interest rate volatility and the term structure: a two-factor general equilibrium model. *Journal of Finance*, **47**, 1259-1282.

[49] Nelson, C.R. and Siegel, A.F. (1987). Parsimonious modeling of yield curves. *Journal of Business*, **60**, 473-489.

[50] McNeil A.J., Frey R. and Embrechts P. (2005). *Quantitative Risk Management*, NJ: Princeton University Press.

[51] Merton, R.C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science*, **4**, 141-183.

[52] Meyer, R. and Yu, J. (2000). BUGS for a Bayesian analysis of stochastic volatility models. *Journal of Econometrics*, **3**, 198-215.

[53] Norris, J.R. (1997). *Markov Chains*, Cambridge University Press.

[54] Nowman, K.B. (1997). Gaussian estimation of single-factor continuous time models of the term structure of interest rates. *Journal of Finance*, **52**, 1695-1706.

[55] Pearson, N. and Sun, T.-S. (1994). An empirical examination of the Cox, Ingersoll and Ross model of the term structure of interest rates using the method of maximum likelihood. *Journal of Finance*, **54**, 929-959.

[56] Pooter, M., Ravazzolo, F. and Dijk, D. (2007). Predicting the Term Structure of Interest Rates: Incorporating parameter uncertainty, model uncertainty and macroeconomic information. *Tinbergen Institute Disccusion Paper*, TI2007-028/4, Erasmus University Rotterdam, and Tinbergen Institute.

[57] Protter, P.E. (2005). *Stochastic Integration and Differential Equations (2nd edn)*, Springer-Verlag: Berlin.

[58] Rogers, L.C.G. (1997). The potential approach to the term-structure of interest rates and foreign exchange rates. *Mathematical Finance*, **7**, 157-164.

[59] Rutkowski, M. (1997). A note on the Flesaker and Hughston model of the term structure of interest rates. *Applied Mathematical Finance*, **4**, 151-163.

[60] Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003). *WinBUGS User Manual*, Version 1.4.

[61] Svensson, L.E.O. (1994). Estimating and interpreting forward interest rates: Sweden 1992-1994. *IMF Working Paper*, **94.144**, 33 pp.

[62] The Faculty of Actuaries and The Institute of Actuaries (2002). Pensioner Mortality Tables, Formulae and Tables for Examinations.

[63] UK Debt Management Office (2005). Formulae for calculating gilt prices from yields (3rd edn).

[64] Vasicek, O. (1977). An equilibrium characterisation of the term structure. *Journal of Financial Economics*, **5**, 177-188.

[65] Wang, J. and Zivot, E. (2005). *Modeling Financial Time Series with S-PLUS*, Springer-Verlag: New York.

[66] Wu, T. (2003). What makes the yield curve move?, *FRBSF Economic Letter*, **15**.